Journal of Educational Research

Official Organ of the National Association of Directors of Educational Research

VOLUME IV
June-December, 1921

PUBLIC SCHOOL PUBLISHING COMPANY
Bloomington, Illinois

Direct.
8-1-1922

Journal of Educational Research

OFFICIAL ORGAN OF THE NATIONAL ASSOCIATION
OF DIRECTORS OF EDUCATIONAL RESEARCH

VOL. IV JUNE, 1921 No. 1

CONTENTS

PUBLISHED FOR THE BUREAU OF EDUCATIONAL RESEARCH · COLLEGE OF EDUCATION
· UNIVERSITY OF ILLINOIS ·

Journal of Educational Researc

Published for the Bureau of Educational Research, University of Illinois

THE JOURNAL OF EDUCATIONAL RESEARCH is published at 450 Ahnaip St., Menasha, Wis., monthly exce[...] July and August by Geo. Banta, Official Printer to the Public School Publishing Co. The subscription price is $4.00; $4.[...] in Canada or Mexico; and $4.50 in other foreign countries.

The editorial office is at Urbana, Illinois. Address all communications on educational matters, articles for the magazi[...] etc. to B. R. Buckingham, University of Illinois, Urbana, Illinois.

The Public School Publishing Company conducts the subscription work. Send all orders to them at Blooming[...] Illinois.

The JOURNAL OF EDUCATIONAL RESEARCH is both needed by and needs the subscriptions of many mo[...] teachers and school officers.

The title page and index for Volume III were furnished with May, 1921 number.

Articles to Appear Soon

The New Plan of Admitting Students at Columbia University by *Edward L. Thorndike*
Analysis of Reading Ability by *S. A. Courtis.*
Measuring Progress by Means of Standardized Tests by *S. S. Brooks.*
Arithmetic Ability of Men in the Army and Children in the Public Schools by *Art[...] Kolstad.*
Correlation: How to Work It on the Adding Machine by *Herbert A. Toops.*
Rate of Progress in Teacher Preparation by *W. Randolph Burgess.*
Method of Equalizing Teachers' Ratings by *Lee Byrne.*

JOURNAL *of* EDUCATIONAL RESEARCH

VOLUME IV | JUNE, 1921 | NUMBER 1

DIAGNOSTIC AND REMEDIAL STEPS IN READING[1]

WILLIAM S. GRAY
The University of Chicago

Any report which is made at this time of diagnostic and remedial work in reading must be tentative and incomplete for two reasons. The technic of diagnosing individual and group needs has not been fully developed. Furthermore, appropriate remedial devices have been organized for only a limited number of disabilities. It is possible, on the other hand, to secure helpful suggestions from a study of specific methods now employed in various laboratories and school systems of the country.

Experience has proved that diagnostic and remedial work in reading should begin with a study of the accomplishments of all the pupils of a school. This step enables teachers to determine the major problems of reading instruction in each grade for the year. It also secures information concerning the needs of every pupil. Unless a systematic program of testing is followed there is danger that many significant individual difficulties will not be noted.

In the Elementary School of the University of Chicago at least four records of accomplishment in reading are secured at the beginning of each school year: (a) mastery of the rudiments, or mechanics of oral reading, including both rate and accuracy, as measured by the Standardized Oral Reading Paragraphs; (b) rate of reading simple material silently as measured by the Courtis' Silent Reading Test, No. 2; (c) ability to understand simple passages as measured by Courtis' Silent Reading Test, or Burgess' Scale for Measuring Ability in Silent Reading (P. S.-1); (d) ability to understand increasingly difficult passages, as measured by Monroe's Silent Reading Test, or Thorndike's Scale Alpha 2. Although these tests do not cover all phases of reading, they secure a

[1] A paper presented at the meeting of the National Association of Directors of Educational Research at Atlantic City, N. J., March 3, 1921.

sufficient amount of information for an exceedingly helpful preliminary diagnosis. In the report which follows the discussions of diagnostic and remedial steps are limited on account of time to one phase of reading.

As soon as the tests have been given and the scores calculated, a careful study is made of the average scores of each grade to determine which phases of reading should be emphasized most during the year. For illustration, Figure 1 represents the average scores in oral reading in October, 1919. The vertical lines are the lines on which the scores for the grades are indicated. The solid oblique line represents the standard scores. The dotted oblique line represents the scores for the University Elementary School in October, 1919.

An analysis of the records revealed two significant facts: (a) oral reading needed little emphasis in each grade as a whole above the III-B, because the average class scores at the beginning of the year were higher than the standard scores for the end of the year; and (b) more or less class instruction in oral reading was necessary in the II-B, II-A and III-B grades because the average class scores were lower than the standard scores.

The second step in the diagnosis included an analysis of the individual scores in each grade. Table I shows the distribution of the pupils of each grade in oral reading. The numbers above the vertical columns indicate the oral reading scores. The grades are indicated in the left-hand column and the total number of pupils in each grade is indicated in the right-hand column. The entries in the table show the number of pupils receiving each score. Thus in the II-B grade, 1 pupil made a score of 5, 1 a score of 10, 3 a score of 20, etc.

The irregular vertical line in the table indicates the scores which pupils should make in oral reading to do most effective work in fluent intelligent reading. In the University Elementary School, these scores are 40 for the II-B grade, 45 for the II-A grade, 50 for the III-B grade, and 55 for all of the higher grades. The scores were determined three years ago after careful comparative studies had been made of the accomplishments of pupils in various phases of reading.

The pupils in the lower grades who ranked higher than these scores were promoted to a higher grade, if their work in other subjects justified promotion. If they were not promoted, they were

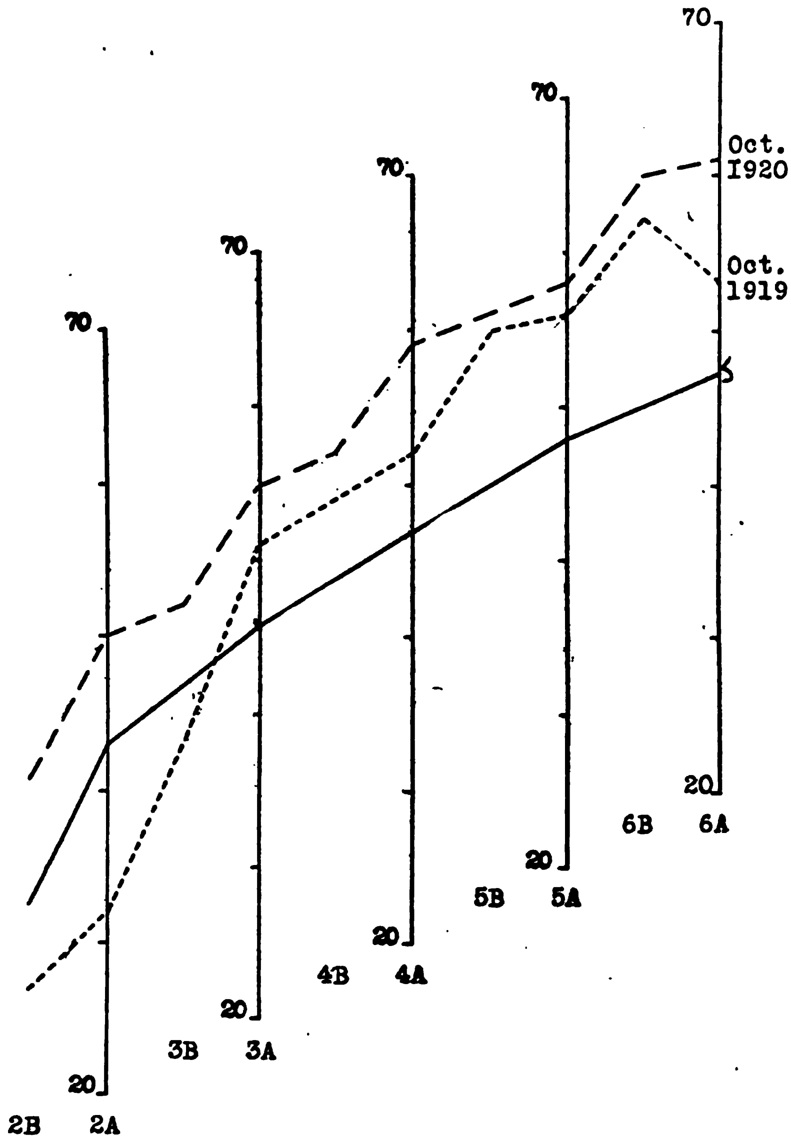


FIGURE 1. AVERAGE SCORES IN ORAL READING IN OCTOBER, 1919, AND OCTOBER, 1920

excused from oral-reading exercises, and either provided with material to read silently or given additional work in those subjects in which they needed help. The pupils who scored lower than the desirable standards were given help in oral reading in proportion to their needs. In attempting to organize appropriate remedial instruction for these pupils, additional diagnostic steps were adopted.

TABLE I. DISTRIBUTION OF PUPILS IN ORAL READING

Grade	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	Total
II-B	1	1		3	1	2	2	3	4	2	5	5		1		30
II-A				1		1		2	3	4	7	4	2	2		26
III-B							3	1	6	8	4	3				25
III-A									3	5	8	4	6		2	28
IV-B								3	8	3	8	3	4		3	32
IV-A										5	5	8	8	3	1	30
V-B								2	1	2	3	7	6	3		24
V-A									1	2	3	9	10	5	1	31
VI-B									1	2	4	9	6	6		28
VI-A										2	4	2	13	5		26

A study was first made of the kinds of errors which pupils make in reading. The most significant errors with words which were discovered are non-recognitions, total mispronunciations, and partial mispronunciations of (a) monosyllabic words and (b) polysyllabic words. The most important errors with groups of words are poor phrasing, omissions, insertions, substitutions, repetitions, and reversing the order of words and phrases.

A careful study was next made by the teachers of the methods of instruction which are appropriate in preventing or eliminating each type of error. Such matters were emphasized as the following: the importance of numerous simple interesting reading exercises

to promote the establishment of effective habits of reading; the value of flash-card exercises in the lower grades in developing a sight vocabulary, and in increasing the span of recognition; the importance of directing attention to the content of what is read in correcting mispronunciations; the value of phonetic analysis in eliminating partial mispronunciations of monosyllabic words; the value of a study of the rules of syllabication and accent in eliminating the mispronunciation of polysyllabic words; the importance of exercises in recognizing thought units in sentences in order to improve phrasing; and the value of careful reading to eliminate omissions, substitutions, insertions, and repetitions.

KINDS OF ERRORS IN ORAL READING

I. With Words
- 1. Non-recognition
- 2. Total mispronunciation
- 3. Partial mispronunciations
    - a. Monosyllabic words
        - Beginnings and endings of words
            - (Usually a consonant or phonogram)
        - Middle parts of words
            - (Usually a vowel or digraph)
        - Enunciation
    - b. Polysyllabic words
        - Syllabication
        - Accent
        - Repetition of one or more parts
        - Substitution
        - Omission or insertion
        - Inaccurate pronunciation of a part or a syllable

II. With Groups of Words
- 1. Poor phrasing
- 2. Omissions, insertions, and substitutions
    - a. Which change meanings
    - b. Which do not change meanings
- 3. Repetitions
    - a. To secure a better attack on a word or phrase
    - b. To correct an error in pronunciation
    - c. To verify meanings
- 4. Reversing the order of words and phrases

The study of devices for overcoming difficulties was accompanied by a detailed analysis of the causes of difficulty in oral reading. Some of the more important causes which have been discovered

appear in the outline. Irregular attendance causes slow progress on the part of some pupils. School physicians and teachers frequently find that poor health, malnutrition, or nervous disorders interfere with rapid progress. Studies of the nationality of the child, his home environment and training, the economic conditions under which he lives and the peculiarities of the section of the country from which he comes have suggested remedial measures, such as a larger amount of training in English to overcome language handicaps, or a generous provision of interesting reading material which may be read at home. Furthermore, studies of the previous instruction of the child have revealed the fact that his difficulties are due to inappropriate or ineffective methods of instruction, to constant use of difficult selections, to an inadequate amount of reading material, or to uninteresting selections which were used.

The organic causes of difficulties in reading are far more difficult to detect and to remedy. Many children experience genuine difficulty because of visual defects which can be remedied through the use of appropriate lenses. Some children find it difficult to form the motor coordinations which are essential in reading. Carefully prepared exercises to strengthen the muscles of the eyes have proved helpful. A limited visual span prevents some readers from recognizing a group of words at a given fixation. This limitation can be detected through the use of a tachistoscope. Short exposure exercises have proved effective in increasing the amount which one can recognize at a single fixation of the eye. A low degree of visual acuity frequently leads to numerous errors in reading. It can be detected through a letter-marking test, and has been improved by exercises of the same type.

Defects in the brain tissues frequently cause difficulties in reading. Word blindness, for illustration, is defined as extreme difficulty in learning to recognize printed or written language by persons of normal mentality and vision. It has been overcome in a limited number of cases by methods of instruction which make a vigorous appeal to the child. Vocal defects, such as malformations of the vocal organs and enlarged tonsils, interfere seriously with effective reading. Breathing irregularities, which are caused by adenoids or by inadequate control of the diaphragm and chest muscles, may lead to errors. Auditory defects, such as word or sound deafness, frequently lead to errors in pronunciation.

A third group of causes are psychological in character. They include ~~general mental incapacity~~, inadequate attention to meaning, failure to associate appropriate meanings with words, limited eye-voice span, limited span of recognition, inability to remember new words easily, capacity to learn words only very slowly forgetting them quickly and easily, and inability to analyze and pronounce words effectively.

With the various causes of difficulty in mind, the teachers made a careful study of the records of their pupils. Those who were below the required standard but who gave evidence of no unusual or peculiar difficulty were given oral-reading instruction appropriate for normal second- and third-grade classes. Special effort was made, however, to interest the pupils and to provide them with a large amount of simple interesting material for use in class and for supplementary outside reading.

The regular instruction which has just been described was supplemented by special exercises to meet the needs of small groups of pupils. For illustration, one group did considerable silent reading to develop habits of thoughtful reading. A second group gave attention to word analysis and phonics to gain independence in word recognition. A third group drilled on quick perception exercises to increase the span of recognition. A fourth group selected thought units in sentences to improve phrasing. Although much time was devoted to these exercises, the pupils were always given abundant opportunity to read in order that the help which was derived from the special exercise might be reflected as soon as possible in their reading habits.

The pupils who did not respond to this treatment were sent to the special "remedial" teacher for a thorough diagnosis and for appropriate remedial exercises. In order to describe concretely the diagnostic and remedial steps which were employed, the case of a fourth-grade boy will be presented in detail, who found it necessary in September, 1920, to give up some of his school work because of inability to read. This situation was so acute that it seemed for a time that he would have to discontinue school altogether. In the diagnosis of his case he was first given the Standardized Oral Reading Test in which he scored distinctly below the standard for the second grade. His record showed that he was unable to recognize many very simple words. He made

numerous substitutions, such as *says*, for *said*, *was* for *were*, *small* for *same*, and *he* for *the boy*. Furthermore, he recognized words individually rather than in groups, which resulted in an unusually slow rate of reading.

The most characteristic difficulty which was revealed was a peculiar form of confusion which led to frequent repetitions. In this connection he frequently pronounced words which appeared to the right before pronouncing those to the left. This error occurred so frequently that detailed consideration was given to it later in the diagnosis.

The Burgess' Scale for Measuring Ability in Silent Reading (P. S. -1) was given next to determine how effectively the subject read silently. In this test he ranked distinctly below the median score for the second grade.

Courtis' Silent Reading Test was given next to secure additional information concerning the subject's rate and comprehension in silent reading. In accuracy of interpretation he excelled the fourth-grade standard, but in rate of reading he ranked very low. The results of this test suggested quite definitely that his major difficulty related to the recognition of words, rather than to their interpretation.

Additional evidence supporting this conclusion was secured through a supplementary test based on passages of the Courtis rate test. After the subject had read three passages silently, he was asked to reproduce what he had read. He reproduced only 16 words, or 12.5 percent of the amount read. He then read three passages orally and reproduced 18 percent. Three passages were then read to him and he reproduced 70.1 percent of the content. The superior score which was made when the subject was relieved of the problem of recognition made it clear that his difficulty in reading related primarily to recognition rather than interpretation. A study was therefore made of the causes of difficulty in recognition. In this connection four steps were taken.

1. Jones' Vocabulary Test was first given to determine the ability of the subject to pronounce words accurately. The records showed that he failed to see many letters unless his attention was called to them, he confused the sounds of important letters of words, he was unable to analyze short words containing the simplest phonetic elements, and was unable to recognize at sight frequently recurring words, such as *what*, *that*, and *you*.

2. The difficulty encountered in the recognition of words suggested the possibility of marked weakness in visual memory. A visual memory exercise was therefore given to the subject and to four fourth-grade pupils who ranked distinctly above the average in general intelligence, school marks, and oral reading accomplishment. The special subject of this study made fewer errors than any other pupils. This indicated that very little if any explanation for slow progress in reading could be attributed to gross defects in visual memory.

3. A study was next made of the ability of the subject to recognize individual letters, parts of words, and groups of words at single fixations of the eye. The materials which were used included the following: the 26 letters of the alphabet, 18 two-letter words, and 10 each of two-letter nonsense syllables, four-letter words, four-letter nonsense syllables, two-word phrases, and three-word sentences. These were presented uniformly through the use of a drop tachistoscope. Each item was presented until it was accurately recognized. The four fourth-grade pupils who took the visual memory test took this series of tests also. Table II contains a summary of the results.

TABLE II. THE NUMBER OF ERRORS IN RECOGNITION BY FIVE SUBJECTS IN THE TACHISTOSCOPIC EXPERIMENT

Recognition Series	Subject A	Subject B	Subject C	Subject D	Subject E
Isolated letters	13	16	15	9	20
Two-letter words	5	6	10	1	2
Two-letter nonsense syllables	7	4	2	1	13
Three-letter words	10	4	4	0	3
Three-letter nonsense syllables	31	9	12	2	6
Four-letter words	19	1	1	2	4
Four-letter nonsense syllables	38	7	17	2	9
Two-word phrases	32	3	4	1	2
Three-word sentences	25	3	16	1	5
Total	180	53	81	19	64

Subject A made more than twice as many errors as any other subject. He recognized isolated letters, two-letter words, and two-letter nonsense syllables as accurately as the other subjects. His difficulties increased markedly with the three-letter units. The

numerous errors made in recognizing three- and four-letter nonsense syllables indicated marked inferiority in the accurate recognition of the details of a group of letters. Furthermore, the errors made in recognizing two- and three-word phrases and sentences indicated that his span of recognition was much narrower than that of the other subjects.

4. Before drawing final conclusions a photographic record was made of the eye-movements of the subject while reading a simple, unfamiliar passage silently. A summary of the difficulties which were discovered follows. (a) The subject recognized each word individually. This is indicated by the fact that there is at least one fixation, and in some cases, several fixations, per word. (b) These fixations were very irregular and unsystematic. This indicates that the subject did not have a definite method of extricating himself from a word difficulty. (c) The subject used an inaccurate, uneconomical return sweep from the end of one line to the beginning of the next. This was indicated by the location of the first fixation of each line. In the second line it was fourth from the left, in the third line it was eighth, in the fourth line it was twelfth, and in the fifth line it was seventh. (d) Instead of beginning at the left and going forward step by step, the eye skipped about, sometimes fixating on a point very much ahead of where it should be and at other times moving to the left over parts which had already been read. A partial explanation for this irregularity is found in a statement made by the subject. He stated that in reading a sentence he tried to find a sufficient number of words which he knew to enable him to guess at the meaning of the rest. This resulted in irregular wandering movements rather than definite progressive movements and made fluent reading more or less impossible.

The diagnosis which has been reported thus far revealed five outstanding defects in the reading habits of the subject: (a) inappropriate motor habits in making the return sweep; (b) irregular progression of attention from left to right; (c) failure or inability to scrutinize words in sufficient detail to recognize significant parts; (d) inability to analyze new words; and (e) inability to recognize words in groups or thought units.

The remedial program of instruction which was adopted included three distinct features. For thirty minutes each day the subject was under the immediate supervision of Miss Delia Kibbe, a special teacher of unusually bright and slow children in the Uni-

versity Elementary School. The problem assigned to Miss Kibbe was the development of ability to recognize words independently. During a second period of thirty minutes the subject was under the direction of the writer. An attempt was made (a) to correct the inappropriate motor habits which were employed in making the return sweep; (b) to promote regular progression of attention from left to right; (c) to increase the ability of the subject to note important details of words; and (d) to increase the span of recognition.

In order to develop effective eye-movements interesting selections were typewritten in three forms. In each type the lines were separated more widely than in ordinary print in order to promote a rapid and accurate return sweep. The subject was instructed to move his eyes quickly from the end of one line to the beginning of the next without stopping to look at any of the words of the second line. After ten five-minute exercises in which this problem was emphasized, he was able to make the return sweep effectively.

The words of the three types of exercises were typewritten in modified form in order to promote regular progression of attention from left to right. In the first type the words were written five-letter spaces apart. The subject was instructed to read the words in order without glancing to the right or to the left of a given word before it was recognized. The fact that the words were separated made it somewhat easier for him to concentrate attention effectively. After ten five-minute exercises of this type he showed marked improvement in regularity of eye-movements and in fluency in reading.

In the second type of exercise the words were grouped together in thought units as far as possible. Exercises of this type are calculated to promote a regular progression of eye-movements, and in addition, the recognition of more than one word at each fixation. If, while reading, the subject gave evidence of wandering movements, his attention was directed to the specific word or words which were causing difficulty. It should be added at this point that during a period of ten exercises the number of corrections which were necessary gradually decreased.

In the third type of exercise the words were written one-letter space apart. The subject read aloud five minutes each day while the experimenter noted for evidences of ineffective return sweeps and irregular progression of attention. If irregularities were

DRILL PASSAGES TO ESTABLISH APPROPRIATE EYE-MOVEMENTS

Type I

Once upon a time there lived in a cottage near a wood a poor widow. In the garden in front of her house there grew two rosebushes, one of which bore white roses and the other red.

Type II

One day when her spindle was so red with blood that the poor girl could not spin, she tried to wash it in the water of the spring; but the spindle fell out of her hand and sank to the bottom.

Type III

These two little girls were the best children in the world. Snow-White was quiet and gentle. She used to stay at home with her mother, help her about the house-work, and read to her after it was done; while Rose-Red liked to run about the fields and look for birds and flowers.

noted, suggestions, calculated to correct the difficulty, were offered. At the end of nine weeks marked improvement had been made as evidenced by greater fluency and fewer irregularities.

A second series of exercises was organized as a means of directing the attention of the subject to important details of words and of increasing the span of recognition. These exercises took the form of eleven drill books each of which contained ten words or phrases of approximately the same length for use in short exposure exercises. The first exercise of each book appears in Table III. The words of Book I were exposed in order. Less than a second was allowed for each exposure. A score of one was allowed for each word accurately recognized. The table shows that on the first day the words of the first three books were recognized with-

TABLE III. RECORD OF THE SHORT EXPOSURE EXERCISES DURING A PERIOD OF TWELVE DAYS

Book	First Exercise of Each Book	Day	Score Made at Each Exposure	Number of Exposures
I	on	1	10	1
II	has	1	10	1
III	bank	1	10	1
IV	of wind	1	6, 8	9
		2	8, 8, 6, 6, 10, 10	
		3	10	
V	in the garden	3	7, 7, 7, 7, 7, 9, 9, 10, 8, 10, 10	12
		4	10	
VI	the willow buds	4	7, 6, 6, 6, 6, 8, 7, 10, 10	17
		5	6, 9, 9, 9, 9, 10, 10	
		6	10	
VII	he said	6	9, 9, 9, 9, 10, 10	7
		7	10	
VIII	pretty soon	7	9, 9, 10, 10	5
		8	10	
IX	What is that?	8	3, 3, 5, 6, 6, 7, 8, 9, 9, 9, 10, 10	13
		9	10	
X	to her fairy story	9	2, 2, 4, 6, 7, 7, 8, 9, 8, 10, 10	19
		10	7, 9, 9, 10, 9, 10, 10	
		11	10	
XI	their little seed boxes	11	3, 3, 2, 5, 5, 6, 7, 10, 9, 10, 10	15
		12	8, 10, 10	
		13	10	

out errors. The phrases of Book IV were exposed twice on that day. During the first trial six of the ten phrases were accurately recognized. The subject was given his score but no statement was made concerning the character of his errors. During the second trial he made a score of 8. On the second day the exposures were continued until the subject had recognized all of the phrases without error twice in succession. On the third day the phrases were exposed again. Inasmuch as the subject made a perfect score on Book IV, Book V was begun.

The procedure which has just been described was continued for twelve days. At the end of that time the tachistoscopic exercises which had been given in the early examination of the subject were repeated. The table includes a summary of the errors made by the subject before and after training on the twelve daily exercises. It shows a decrease in the total number of errors from 180 to 122. There was a decrease in non-recognition from 39 to 32, and in wrong recognitions from 141 to 90. An examination of the records for each type of material used shows satisfactory improvement in the recognition of isolated letters, three-letter words, three-letter nonsense syllables, four-letter words, two-word

TABLE IV. THE NUMBER OF ERRORS IN RECOGNITION BY SUBJECT "A" IN THE TACHISTOSCOPIC EXPERIMENT BEFORE AND AFTER TWELVE SHORT EXPOSURE EXERCISES

Recognition Series	No.	Number of Errors Before Training			Number of Errors After Training		
		Non-recognition	Wrong Recognition	Total	Non-Recognition	Wrong Recognition	Total
Isolated letters	26	6	7	13	4	0	4
Two-letter words	18	0	5	5	3	2	5
Two-letter nonsense syllables	10	1	6	7	4	2	6
Three-letter words	10	0	10	10	2	2	4
Three-letter nonsense syllables	10	4	27	31	3	12	15
Four-letter words	10	4	15	19	1	1	2
Four-letter nonsense syllables	10	8	30	38	13	42	55
Two-word phrases	10	6	26	32	0	15	15
Three-word sentences	10	10	15	25	2	14	16
Total		39	141	180	32	90	122

phrases and three-word sentences. For some unexplained reason the subject did less well on four-letter nonsense syllables than in the original test. Inasmuch as the number of errors made in the recognition of several of the exercises was distinctly above the average made by effective readers in the first test, it was evident that short exposure exercises could be continued to advantage.

After two months of remedial work, the subject was thoroughly tested again to determine progress. In the oral reading test, he made a score of 37.75. This indicates that he made approximately a year's growth between October and December.

In the Burgess test, he made a score of 50 on the third-grade scale and a score of 34 on the fourth-grade scale. The satisfactory improvement may be attributed almost wholly to increased rate of reading.

In the Courtis' Silent Reading Test the subject answered twenty questions which is twice the number answered in the first test. He also made a comprehension score of 90 which is equal to the score in the first test.

The December tests showed clearly that the remedial work had resulted in greater fluency and effectiveness in reading. Although the subject was not fully up to standard by the middle of the current school year, he was permitted to resume regular work with the fourth-grade class, and he promises at the present time to make a very creditable record. If he had not been subjected to individual diagnosis, it is probable that the real nature of his difficulty would not have been discovered. Too many pupils fall into the retarded group, and finally drop out of school, because of difficulties which could be removed. In my judgment, each school system of considerable size should establish a center to which unusual cases could be taken for diagnosis and remedial instruction.

THE NATIONAL INTELLIGENCE TESTS[1]

Guy M. Whipple
Bureau of Tests and Measurements, University of Michigan

Historical Statement

Before the war both Professor Yerkes and Professor Terman approached the General Education Board for the support of a sort of school survey which would include the measurement of the intelligence of a good-sized group of pupils. The success of the Army Alpha Intelligence Examination made it evident that the same general methods would be applicable for such an examination of intelligence and that there would almost certainly be attempts made on the part of various individuals who had had contact with the army methods to adapt these to the examination of school children. It was felt that it would be very advantageous to the whole movement of mental testing if this adaptation could be made carefully, systematically, under the auspices of some institution or organization with prestige, and by men who would make a serious and expert contribution. The General Education Board acted favorably upon these suggestions with the proviso that the National Research Council should take the responsibility for the undertaking and that a group of four or five psychologists should cooperate in working out the details. A sum of money was appropriated for the work, and Messrs. Haggerty, Terman, Thorndike, Yerkes, and the speaker were made the members of the Committee.

We met first at Washington March 28–29, 1919, again April 29 to May 2, 1919, a third time October 17–18, 1919, and at Chicago in December, 1920. A preliminary printing of trial tests was made in the spring of 1919, and the final completed scales were issued in the summer of 1920. Something like 200,000 copies have been sold to date. The authors' royalties, I may add, are turned over to the National Research Council for use in further studies of tests.

It is my object in a few minutes to say something about the aims of our committee, the methods by which its work was

[1] Paper read at the meeting of the National Association of Directors of Educational Research, at Atlantic City, March 3, 1921.

done, the criteria we adopted in selecting and arranging the scales, and the results which are being obtained.

Scope

Our aim was to produce in a single pamphlet an arrangement of tests that could be applied to any child in the elementary school who could read well enough to participate in a group examination. In practice this means from the upper half of the third through the eighth grade. An examination that serves satisfactorily in the ninth grade or above is almost certain to be too difficult in content or in time limits to suit pupils in the lower grades. The National examination has been given with some success in the lower half of the third grade and, though the distribution tends to be "bunched" toward the lower end, yet even in that grade it operates well to locate the brighter pupils and does not give over 2 percent of zero scores. (Six in 363 were reported at Jackson, Michigan.)

Range of Difficulty

In fitting the individual tests to this proposed scope of grades and ages there was, necessarily, a problem of selecting material that would afford a proper range of difficulty. The criteria provisionally adopted in this respect, which were, I believe, successfully met with practically every test, were that *each test* should be so contrived that not more than 10 percent of perfect scores should be made by the average group of eighth-grade pupils and not more than 10 percent of zero scores should be made by the average group of upper third-grade pupils. I have not had time to verify the operation of these criteria with scores from various school systems, but I am sure that they are met, or nearly met, under usual conditions of testing.

Progressive Difficulty of Items

In order to produce a single examination that could be used both in the upper third grade and in the eighth grade and that would avoid too many zero scores in the third grade and too many perfect scores in the eighth grade, it was felt by all members of our committee that the several items within each test should vary in difficulty and should be arranged approximately in order from least difficult to most difficult.

I am aware that there are theoretical arguments against measuring speed and degree of difficulty at the same time, and that the arrangement just mentioned may be claimed to be an attempt of that sort.[3] Nevertheless, the experience of the Alpha Army Test lends support to our position; and it may be argued that if the difficulty of the items is properly distributed and the time-limit is properly set, the arrangement we have made is defensible. In any event, while we cannot argue that the steps of our scale are everywhere equal steps (while, for instance, a score of 120 may not be just 20 percent better than a score of 100), nevertheless, we can argue that the higher the score, the better the intelligence of the examinee, so that, for children of a given grade or of a given age, the total scores in either Scale A or Scale B can be regarded as indicating relative orders of merit and hence can be taken directly or indirectly as bases for classification, which is, after all, the final test of usefulness.

It remains to be said on this point that the actual arrangement of the items in the National Intelligence Tests is an empirical one; it corresponds to the order indicated by actual test results.

Coaching and its Prevention

At the first meeting of our committee there was much discussion of the danger of coaching and of the means of meeting it. Professor Thorndike, in particular, urged consideration of this point. He felt that not only would intelligence testing become a common feature of public school administration, but that it would become also a common feature of business administration, e. g., that business men would use mental tests in selecting young boys and girls for beginners in their establishments. He argued also that within some five years practically every city of over 25,000 population would establish special classes for gifted pupils and that many parents would seek to coach their children to pass intelligence examinations given for the selection of pupils for these classes. On this account, he urged that any intelligence examination that came into general usage ought to be capable of almost unlimited expansion, that it was desirable to include in our examinations only tests the material of which could be so

[3] The recently issued Ayres reading test, for instance, is based on the principle that in a rate test all items should be of equal difficulty and rate should be measured directly by quantity done within a given time limit.

extensive that it would be possible to produce even as many as thirty or forty different forms of the test just by drawing material by chance from the general reservoir of items prearranged for each test—this, even though the actual forms might not be as good as could have been produced by selecting the best of the items and confining attention to the production of, say, four or five of the best possible forms. He further argued that another way to discourage coaching would be to dump a considerable number of forms upon the market at the outset—enough to discourage any one who tried to anticipate and prepare for the examination.

The other members of our committee felt that the danger from coaching was much less serious than this proposal would imply, and there was argument for restriction of our efforts to a few forms. The compromise finally effected seems to me to meet the situation adequately. To begin with, the National Intelligence Tests appear in two independent scales, A and B, either one of which will serve well for an intelligence examination. Second, there is to be published very soon a second form of each scale (the proof is even now in hand), and this will be followed at intervals by three more forms, until five forms of each scale are available. These five forms, as will be shown later, have actually been tried with school children under proper conditions to cancel out errors of time-order. Then the forms have been edited, very slightly, altering an item here and an item there, to bring them to what is probably as near an exact equivalence as can be produced in instruments of measurement of this sort. I shall return to this point in just a moment. Third, there has been deposited at the offices of the National Research Council at Washington, complete material for five additional forms (that is, 10 in all) for each test in each scale. The material for Forms 6 to 10 was drawn by lot from material gathered by the several experimenters, just as was the material for Forms 1 to 5; there is every reason to expect that without any further trial it will afford a set of tests as nearly like the first five forms as these are like each other. Fourth, reports have been filed at Washington showing in detail the methods used by each member of the committee in preparing the items of the tests for which he was responsible. This will make it possible, if the need ever arises, to carry these methods forward and produce still other forms.

As the matter lies, then, our committee has already prepared and tested in actual use five complete forms of each scale, so that if we suppose that our five experimentally equalized forms of each scale are in print and that a prospective examinee should really sit down to prepare himself in advance to cope with any form that might be put before him, he would have to drill himself on the following:

1. 80 examples in arithmetic
2. 100 sentence completions
3. 120 logical selections
4. 200 synonym-antonyms
5. 45 symbol-digit associations (which conflict with one another)
6. 110 arithmetical computations
7. 200 pieces of information
8. 200 vocabulary items
9. 160 analogies
10. 250 same-or-different comparisons

If he succeeds in preparing these 1,865 items, "I'll say" for one that he has had a liberal education and deserves to "pass" the test.

Incidentally, it may be noted that the completion of the 10 forms necessitated the assembling, testing, and arranging in proper form and order of difficulty of 3,730 individual items, which may explain in some measure why five grown men used up nearly twenty-five thousand dollars worth of time and supplies in producing these scales—about six or seven dollars per item—a matter that always seems to puzzle the amateur mental tester and the layman who thinks that a mental test is something that can be thrown together overnight by any one with a high-school education.

Examiners

The aim of the committee as originally defined was to produce an intelligence examination, to quote the resolution adopted, "that can be used by intelligent normal school and college graduates with a reasonable amount of special training." As a matter of fact, the effort was made, however, to render the administration of the examination even easier than this statement would imply. And experience has shown that almost all classroom teachers in the elementary school can be brought, within a brief time, to administer the examination in proper style. Of

course, mistakes occur, and in any large school system there is usually one room from which the results obtained point to some error in applying the scales. This is inevitable when any sort of standardized procedure is attempted and need not be charged against the National Intelligence Tests as a fault of design. Mistakes of this sort are usually self-revealing and can be corrected by a second trial with another form of the same scale.

In Michigan cities where wholesale intelligence examining has been attempted under the direction of a central administrative officer, it has been found feasible to utilize the classroom teachers as examiners by distributing copies of the Manual of Directions in advance, with assignments of reading covering the administration of the scale to be used, then bringing the teachers together for consultation, giving them the test (with shortened time-limits), soliciting questions about details, and emphasizing certain points (like manner of giving the directions, strict adherence to the directions, careful timing, etc.) that are most likely to be overlooked or misunderstood.

Total Duration and Timing

A group intelligence examination for use in the elementary schools ought, if possible, to be capable of administration within a single period of forty minutes. Either Scale A or Scale B of the National examination can be concluded within that limit.

In the distribution of time within that limit, our committee proceeded on the assumption that a variety of comparatively short tests was psychologically preferable to a limited number, say two or three tests, of longer duration. For this reason it was agreed that no one test should exceed five minutes and that the time should, as a rule, be placed at two or three minutes without ever using fractions of a minute (this, of course, to lessen errors in timing). The final time limits are, for the five tests in Scale A: 5, 4, 3, 2 and 3 minutes, and for the five tests in Scale B: 4, 4, 3, 3 and 2 minutes, respectively.

Responses

Another principle which is essential to the success of group tests of intelligence, at least when working with pupils in the elementary school, is that responses should entail a minimum of writing, should be readily understood by the examinee, and

should be unambiguous to the scorer. Some of the group tests that have recently flooded the market take no account of these principles, which have been found in practice worthy of heed and which have contributed definitely to the usefulness of the National Intelligence Tests.

Scoring

To make a group test practically usable, the scoring must be rapid, objective, unambiguous, and simple enough to be undertaken by any clerk of reasonable intelligence. Here, again, I believe that the National examination, with the possible exception of the completion test (I say "possible" exception because even here we supply a fairly complete key) is superior to many other group tests of intelligence. Experience shows that no test can be scored without error, when the task is to handle papers in a wholesale fashion. The commonest mistakes made by unskilled scorers are (1) subtracting wrongs from attempts instead of wrongs from rights in the R – W form of scoring; (2) giving one-half as many or twice too many credits in logical selection; (3) failure to multiply loaded scores; (4) adding wrongly in computing the final score; (5) mistakes in counting up scores on a single test. The experience of Michigan users suggests that, when used in quantities, the tests should first be scored by the teachers, and that all papers should then be shipped to the office of the chief examiner and checked or rescored by clerks trained for this work. If the first half of the papers returned by a given teacher are found to be without error, the remainder may be regarded as sufficiently certain to be without systematic errors, and the work of the clerk may be limited to checking the copying of the figures and the addition for the final score.

The National Intelligence Tests are open to a small element of criticism in two points. First, the symbol-digit test (Scale A, Test 5) necessitates multiplying by 0.3. Inaccurate teachers are apt to multiply by 3, but this mistake is readily detected on rescoring. Further, the decimal that results in this test, to which some object, can be avoided by taking the nearest integer for the score. Second, the method prescribed for scoring the logical selection test when three responses were marked of which two were correct proved bothersome to many teachers. This particular

rule was formulated, I may say, over my objection (I had charge of the test in question).[3]

The keys that are provided for use in scoring have, so far as I know, met with universal approbation and contributed materially to the successful use of the examination.

INSTRUCTIONS

Unusual pains were taken by our committee with the preparation of the instructions for each test. Special endeavor was made to obtain simplicity, brevity, concreteness, and clarity, with emphasis upon clarity.

When clarity could not be obtained with brevity, there was no hesitation in sacrificing brevity. Thus, the instructions for the logical selection test are somewhat lengthy, and those for the symbol-digit test still more so, but this length was decided upon by actual trials that showed the necessity for every word used.

A good example of concreteness is afforded by the directions for the analogies test, which differ, for instance, from those commonly given to adults in that all reference to the term "relation" or the term "proportion" is avoided and dependence is laid upon concrete directions for conveying the idea of the task.

It is my belief that the National Intelligence Tests can claim unusual merit in the formulation of the test instructions, and that the success attained in this regard is another good example of the necessity, in the preparation of tests for children, of using nothing that will not stand the acid test of experience under working conditions.

The essence of the instructions for each test is not only definitely brought out in a fore-exercise; it is also printed concisely at the head of the test proper, where it can be consulted if need be.

FORE-EXERCISES

Our committee decided at the outset that each test should be preceded by a suitable fore-exercise that should consume not to

[3] I have recently checked over enough papers to prove that the simpler method of scoring that I urged at the outset is perfectly satisfactory and will not appreciably alter the scores or standards for the entire examination, except to reduce two or three points the average scores for pupils in the third and fourth grade. The forthcoming supplement to the Manual of Directions for these tests will remove the unnecessarily complicated method of scoring that was originally prescribed for this particular test.

exceed one minute and in most cases not to exceed a half minute, that should be so designed as to make still clearer the nature of the task to be performed, and that should at the same time serve to equalize the knowledge with which the examinees undertake the test in case some of them may have had previous acquaintance with the test and others not.

This feature of the National Intelligence Tests has met with universal approbation and is certainly one of their points of merit.

Another feature of the instructions of the tests is the introduction, in the fore-exercise and at the head of the tests themselves, of samples in which the task is set forth and the method of responding is made clear. This feature is, of course, not peculiar to these tests. Its usefulness is apparent without argument.

Choice of Tests

At the first meeting consideration was given to practically all varieties of tests that had been proposed for use in group examinations. The experience with the army tests was canvassed, and each member of the committee contributed suggestions at length. Various tests were rejected by unanimous consent, e. g., memory for digits, the maze test of Beta 1, the counting-cubes test of Beta 2, etc.

Agreement was finally reached upon a list of twenty-two tests that seemed worthy of preliminary trial. These tests were the following:

Printed directions	Geometrical construction, or form combinations
Disarranged sentences	Copying designs
Arithmetical problems	Vocabulary
Information	Picture sequence
Opposites	Pictorial analogies
Practical judgment	Recognitive pictorial memory
Number series	Sentence completion
Analogies	Pictorial similarities
Series completion	Computation
Symbol-digit	Logical selection
Comparison	
Picture completion	

There was discussion, also, of the possibility of developing an "omnibus" test and of a test that would entail ability to organize, to carry through material demanding a wider scope of attention than necessitated by the "response" type of test illustrated in the twenty-two tests just enumerated.

The Preliminary Series

These tests[4] were prepared by various members of the committee and finally put forth in printed form in eight booklets.

Verbal A comprised six tests (arithmetical problems, directions, information, opposites, practical judgment, and analogies)

Verbal B comprised five tests (computation, vocabulary, sentence completion, disarranged sentences, and logical selection)

Non-Verbal A comprised five tests (picture completion, series completion, number comparison, symbol-digit, and form combinations)

Non-Verbal B comprised five tests (picture absurdities, copying designs, picture sequence, recognitive memory, and pictorial similarities).

The fore-exercises for each booklet were in these trials assembled together in four additional booklets—a plan which was later abandoned in favor of the insertion of each fore-exercise just in front of its test.

These four booklets of twenty-one tests were applied to pupils in the public schools of Alexandria, Virginia, Richmond, Virginia, in the Horace Mann School, New York City, and in three private schools in Cleveland, Ohio. The data from these pupils were forwarded to Dr. Truman L. Kelley, at Columbia University, along with teachers' estimates of intelligence, information concerning the age and grade standing of each pupil. Dr. Kelley, assisted by a staff of clerical workers, subjected the data to very elaborate study. In the fall his report was used by the committee to supplement its data gained in other ways concerning the merits of the several tests.

The result, without going into details, was a decision to utilize ten of the twenty-one tests that had been tried and to issue these ten tests in two batteries of five tests each, either battery as good as the other and each composed of an entirely independent array of tests.

The merit of this decision seems to me unquestioned. With five forms available soon in either scale (and ten if need be), the examiner can repeat a given battery with another form and secure virtual identity of tests and equivalence of items and score or

[4] Number series, though in the list of twenty-two tests, was not tried in these pamphlets.

he can repeat with the other battery and thus get a measurement of the intelligence of his group from a different, but equally valid angle of fire.

Non-Verbal Tests

In the preliminary series of tests a definite effort was made, as I have indicated, to devise non-verbal tests that would show satisfactory validity as measures of general intelligence—this because the committee felt, as most experimenters have felt, that there is danger of stressing linguistic ability to the point of identifying it with general intelligence.

The outcome of this attempt was disappointing. Probably some of the non-verbal tests that we tried could be further improved,[5] but we could not find more than two non-verbal tests that seemed to justify their retention, and these two have been used, one in Scale A and one in Scale B.

Power Tests

The term "power tests" has been applied, perhaps somewhat loosely, to tests in which there is no time-limit or in which the time-limit is so extended that most examinees go as far through the series of progressively more difficult items as their ability permits before the signal to stop is given. Many persons feel that there are certain forms, at least, of test in which speed should be subordinated to "power," in the sense just described. In the National Intelligence Tests, Nos. 2 and 3 in both Scale A and Scale B are planned to operate as power tests. Examiners should not be surprised, therefore, as some have been, to find children finishing their work with these tests before time is called.

Equalization of the Five Forms

I spoke a moment ago about the trial of the five forms of each scale and the minor adjustments that have been recently made to accomplish their practically complete equalization. You may be interested to know how these forms came out in this trial.

These forms were applied to six hundred pupils in grades high III to high VIII, inclusive, in certain schools at Washington, D. C. Without stopping for the scores in the individual tests,

[5] I still think, for instance, that the form combinations test upon which I experimented, has possibilities that a better form of presentation might bring out.

the total scores for the several forms before any attempt at adjustment was made, were as follows:

	Form				
	1	2	3	4	5
Scale A.....	114.9	111.8	112.3	114.4	113.6
Scale B.....	108.4	109.1	112.0	108.6	109.5

These figures furnish excellent evidence of the reliability of the method by which the several forms were prepared. The minor adjustments of which I spoke may be assumed to bring the forms to a degree of equivalence such that the total score of any form will not, at the outside, differ from the total score of any other form of the same scale by more than 0.5 points, or about 0.5 percent.

Results Obtained

The norms in the Manual.—In the first edition of the Manual of Directions tentative norms were supplied, based upon results from 2,000 pupils in Washington, D. C. and 2,000 pupils in Pittsburgh, Pennsylvania. These norms apply to pupils tested about June 1 and tested with Scale B on the day following the test with Scale A. These norms were confessedly tentative.

The results we have obtained in various Michigan cities run almost uniformly below those reported in the Manual. On the other hand, norms recently reported from East Orange, New Jersey, and those reported from certain California cities run higher than those reported in the Manual.

In comparing these results several things have become obvious:

1. Communities differ decidedly in the general level of intelligence as revealed by the tests. The results we have on file at the Bureau of Tests and Measurements at Ann Arbor leave no doubt upon this point. The figures for various Michigan cities, for instance, differ consistently from one another, grade compared with grade and age compared with age. We have concluded that each city must work out its own standards, and that probably it will be worth while eventually to work out certain state standards.

2. The norms reported from various communities are definitely affected by the composition of the groups under test. Thus,

some communities do, some do not, include pupils in speed classes. Some include negro pupils; some exclude them. Some begin at the lower third grade; others at the fourth grade. Some go only through the sixth; others through the eighth. Some have gone above these limiting grades to piece out their age standards by including pupils advanced in the grades; others have not. It follows that age standards (and grade standards, too) will have significance for comparative purposes only when the limitations of the groups under test are definitely stated and definitely allowed for.

3. The time of year at which the examination is held is a more important factor than seemed at first probable. Two graduate students in the mathematics department at Michigan volunteered to derive a formula for correcting scores to a uniform date. This work has been done, and we shall include the corrective formula in the next edition or in a supplementary edition of the Manual of Directions. Incidentally, it may be noted that these corrections are complicated by the varying ages of the groups that enter the schools in September and in February and by the proportion of repeaters in various grade groups. The correction for Scale A is 1.36 points per month. In addition, the correction for a class entering in September when compared with one entering in February is plus or minus 2.0 points, depending on which way the comparison is to be made.

Deriving mental ages and percentiles.—In illustration of the point just made—that city school systems will need to derive their own standards—I may cite the method adopted for this purpose in Jackson, Michigan.

The National Intelligence Tests, Scale A, were applied in that city to about twenty-five hundred pupils in grades III to VI, inclusive. Pupils aged ten and eleven years in the seventh grade were also assembled and tested to perfect the distributions for those ages.

Since most of the information concerning the location of children in the grades is familiar to teachers and supervisors in terms of mental age, it was felt worth while to translate the scores of the National tests into "Jackson mental ages." This was accomplished by regarding the median score of pupils of each age group as the standard score for the mental age as well as the chronological age of the group in question. Thus, all pupils aged

eight (over eighth birthday and under ninth birthday) were distributed in such a way as to locate the median and all the other deciles, and this median was regarded as indicating a mental age of 8½ years. The medians for 9½, 10½, and 11½ years were located similarly and points midway between these medians were taken as the scores indicative of mental ages of exactly 9, 10, and 11 years. The amount of overlapping was shown graphically by the percentile chart, and this chart became directly useful in locating pupils of any desired degree of deviation from the standard adopted for a given grade or group. Thus, pupils were drawn off for consideration in connection with special classes and speed classes, for double promotions, etc.[6] The chart of percentiles shown as Figure 1 is contributed by Miss Helen Davis, one of the recently elected members of this association.

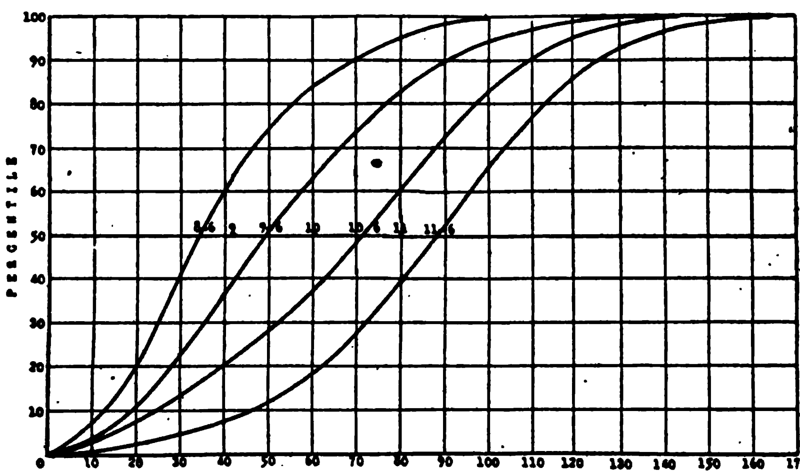


FIGURE 1. PERCENTILE CHART FOR THE NATIONAL INTELLIGENCE TEST, SCALE A, FORM I. (DATA FROM JACKSON, MICHIGAN)

Relation to Binet results.—Another possible method of determining the mental age equivalents of the National examination

[6] It will be understood that in this chart each of the four age-groups of pupils has been reduced to a theoretical 100 pupils. The figures on the base line are the scores obtained; the figures on the vertical lines are the numbers of pupils in order of excellence. Thus, in the group aged 9 years (median age approximately 9 years, 6 months) the twentieth pupil in a hundred counting from the poorest pupil scores 28, the fiftieth (or median) pupil scores 49, the eightieth pupils scores 87, etc. Or, again, 25 percent of the 8:6 group score as high as the median of the 9:6 group, etc.

scores is, of course, to equate these scores with the results of Binet examinations. I am not sure that this method is theoretically as justifiable as the one just described, namely, that of working directly to the mental age by reference to the median score. In any event, I do not have available at this moment statistics to indicate the equivalence to mental ages in the two systems of testing.

I have some figures of interest that refer to a group of thirty-two unusual cases from the Jackson, Michigan, groups just mentioned. Twenty-four of these are pupils of inferior intelligence, under consideration for transfer to an ungraded class or for withholding of promotion. Their I. Q.'s range from 58 to 88, as tested by the Stanford Binet. Their mental ages, as located by the method of determining mental ages for the National tests previously described are in 12 cases from 1 to 18 months lower, and in 12 cases from 3 to 22 months higher than the Binet mental ages: on the average, the National mental ages are 15 days lower, so that the net agreement is striking.

Five of the cases are pupils of superior intelligence, under consideration for transfer to a speed room or for extra promotion. Their I. Q.'s by the Stanford Binet range from 112 to 161. Their mental ages by the National tests are 5 and 28 months lower in two cases, and 5, 7, and 8 months higher in the other three cases. Here the National tests give a lower rating by several months on the average. In view of the fact that these mental ages are mental ages as determined by the local results in the city of Jackson, where the general scores by ages run lower than those reported in some cities, I am at a loss to account for the seeming discrepancy, though the instances are too few to be significant.

Finally, there were in this group three special cases of interest. The first was a speech defective whose Binet I. Q. was 79, but whose National Intelligence rating ran 24 months lower than the Binet mental age. The second was a pupil with an I. Q. of 90, described as a very poor reader, whose National mental age was 13 months lower than that assigned by the Binet test. The third case was a pupil whose Binet I. Q. was 112, but who is likewise characterized as unusually poor in reading. Here, again, the National mental age is strikingly discordant, being 38 months lower than that assigned by Binet testing.[7]

[7] The Binet mental age was 10 years, 8 months. The National score was only 21, or roughly that of pupils about 7½ years old.

These three cases strongly suggest, what would be *a priori* intelligible, that pupils who have special difficulty in reading will suffer a decided reduction in mental age rating by a group intelligence test when compared with their rating by the individual and oral examination of the Binet, though it remains possible that the former rating may be the more significant in predicting school progress.

They suggest, also, that pupils whose scores do not accord with what is anticipated by teachers or who are known to be possessed of some special ability or disability ought probably to be given individual examinations before any important alteration is made in their school status.

Conclusion

Speaking as a member of the Committee who devised the National Intelligence Tests, I presume I am influenced by a certain amount of unconscious bias in their favor. I have, however, sought to speak objectively, to present as directly as I could, some of the facts in the brief history of these tests, the aims of their makers, the criteria that were observed, and the results that are being obtained. If I may be permitted, therefore, to let my sentiments dictate a final sentence, I should like to say that I feel that the committee has a right to feel a tinge of pride in what it has accomplished.

THE MEASUREMENT OF LANGUAGE: WHAT IS MEASURED AND ITS SIGNIFICANCE[1]

ERNEST J. ASHBAUGH
Director Bureau of Educational Service
Extension Division, State University of Iowa

If we include under language all forms of verbal expression, we have to consider not only the most extensive of all the possible fields of pedagogical measurement but also the most complex. We have now progressed far enough in the measurement movement to recognize the fact that a single measure of any complex is quite as likely to conceal as to reveal, and that it is of very little value from the standpoint of diagnosis and remedial work.

This paper confines itself to the field of written expression. It is realized that many other phases might well be discussed under the term "measurement of language," but a desire to keep the paper within the probable time allotment, and to speak very definitely rather than in a general way, forces this limitation.

The various scales for the measurement of composition illustrate very well the above statements. The Hillegas Composition Scale, the Thorndike Extension of the Hillegas Scale, and the Nassau County Supplement by Trabue were all designed to measure general merit with the consequent result that the scales are very difficult to use and the result obtained does not greatly assist the child or the teacher in remedying defects and in thus progressing toward the goal desired. It is to be remembered that the case under discussion is the measurement of language, and what is said here and will be said later pertains directly to the use of these scales as measuring instruments.

Ballou in developing the Harvard-Newton Scale attempted at least partially to remedy this defect by constructing separate scales for each of the four forms of composition. However, any one who has read many children's themes will instantly recall that children seldom write themes which are wholly narrative, descriptive, argumentative, or expository. They have a very disconcerting way of mixing two or more forms into one inglorious whole.

[1] A paper presented at the meeting of the National Association of Directors of Educational Research at Atlantic City, N. J., March 3, 1921.

The latter scale does attempt to be of assistance to the teacher or supervisor in one way in that certain merits and demerits of each of the samples in the scale are pointed out. This is certainly a commendatory feature, but the scoring of compositions for general merit even with these aids is still far from an objective matter. Only careful intensive training will develop one to the point where he can be fairly sure his scoring will have slight deviation from the norm or established value.

The composition scale developed by Breed and Frostic for the sixth-grade attempts a narrower range of merit with consequently, a greater homogeneity, but in other respects it adds little or nothing to the solution of the problem beyond that accomplished by the scales previously named.

Willing made a distinct contribution when he provided in his scale for separating the mechanical elements of capitalization, punctuation, spelling, and paragraphing from the thought element. After having done this, however, he seemed to have felt under the necessity of combining for a single score; and thus, so far as the score itself is concerned, it is no more enlightening than it would have been if he had not separated the two factors in the first place. The accuracy with which the score is secured is probably greater; but whether the compostion is poor because of mechanical defects or because of weakness in thought elements is completely hidden.

What has just been said concerning the scales for measuring composition is not to be construed as an argument against their use. A trained person using a scale can doubtless obtain valuable information concerning the general merit of compositions written by a given class; and this information may be rendered still more valuable when compared with the norm for children of the same grade elsewhere. There is some evidence that scales designed to measure general merit may be used rather effectively as a teaching device in much the same way that the various handwriting scales have proved to be valuable. When samples from the scale are placed upon the wall of the room and the children are encouraged to compare their own products with them, they seem to be able to study these samples and modify the quality of the compositions in terms of the general qualities considered in the scale.

What has been said is only designed to point out the nature of the product which is to be measured and the fact that its com-

plexity makes the measuring instrument of very little value from the standpoint of diagnosis and remedial work.

Any analysis of written expression (which of necessity must be the phase of language ability which will be most frequently measured) must recognize at least three groups of factors, namely mechanical, grammatical, and rhetorical. Each of these groups contains many factors or elements, and only as we separate these complexes into simpler and simpler elements will our measurement become truly helpful to the teacher and supervisor in the improvement of the work in the classroom.

Trabue in his Completion Test Language Scales attempted to devise an instrument "for the measurement of ability along certain lines closely related to language." He states that "Ebbinghaus who invented the completion test method characterized it as 'a real test of intelligence.'" Those who have used these scales have been more inclined to characterize them as intelligence tests than as tests of language ability. Using Scale A Trabue reported a correlation with Hillegas Composition of 0.72 for a group of 30 seventh-grade children, a correlation of 0.85 with the Binet tests for a group of 39 boys, and a correlation of 0.74 with the Binet tests for a group of 50 boys and girls. Dr. H. A. Greene of the University of Iowa, using Trabue's Scale B and Hillegas composition scores from 132 high-school students in one school, found a correlation of only 0.38. In another high school the same tests gave a correlation of 0.14 for a group of 58 students.

From these data it would seem that native ability rather than school training is the determining factor in accomplishment on this test. Possibly the same is true, though to a lesser degree, in achievement in writing compositions. To the extent that scales measure intelligence rather than training, their use in the solution of the teacher's problems in language instruction must be negative and general rather than positive and specific.

Greene's Organization Tests are the result of a definite attempt to analyze language ability and to present an instrument for the measurement of a single phase of this very complex problem. He does not claim to have reduced the problem to the measurement of a simple element. The ability to take a limited number of ideas presented in jumbled order and to rearrange them in logical order is evidently closely related to language ability and is certainly simpler than the ability involved in writing a composition.

He found a correlation of 0.52 with composition for a group of 135 high-school pupils and a correlation of 0.47 with composition for a group of 109 elementary-school pupils. Since his correlations with Terman's mental age (individual measurements) were 0.41 and 0.50 respectively with high- and elementary-school pupils, he concludes that his test is a better measure of the elements involved in language ability than is Trabue's Completion Test. In a recent article in the JOURNAL OF EDUCATIONAL RESEARCH Brooks stated that this test is "mostly an intelligence test" but failed to present any evidence in support of his statement.

Granted that this test will measure somewhat accurately the phase of language ability which it is intended to measure, the fact remains (with this test as with many others) that teacher and supervisor remain quite as much in the dark after as before its use concerning what should be done to improve this ability. A knowledge of low scores is not especially helpful unless at the same time remedial treatment is at least suggested; and thus far, practically nothing is known concerning the type of teaching or drill work which will improve language ability and result in greater achievement on a later test.

Starch, in his Grammatical Scale A, limited himself to the factor of grammatical errors. The scale presents two forms of expression, one correct and the other incorrect, and the child is to discriminate between them, crossing out the incorrect form. The task is thus very definite, the distinguishing between correct and incorrect grammatical forms. The sentences are grouped in fours. These groups are called steps, and the successive steps are supposedly of equal increments of difficulty. The scoring is in terms of these steps of unit value.

Rather extensive use of the scale has brought out the following facts:

1. The test shows rather definitely the sentences in which the pupil cannot distinguish between correct and incorrect form and thus furnishes the teacher with a clue to the instruction needed.

2. The ability to recognize correct form seems to be more closely related to language habits than to knowledge of technical grammar.

3. Intergrade differences are so small when using Starch's method of scoring that they become insignificant and hence the score means little.

4. On the basis of returns from children in grades seven and eight, there seem to be rather wide differences in the difficulty of individual sentences within the same group. (See Table I.) Hence, some other scheme of scoring would probably be of greater value. The facts given in Table I represent a larger number of

TABLE I. STARCH'S GRAMMATICAL SCALE

Step	Sentence	Accuracies (Percent)	
		Grade VII	Grade VIII
5	1	85.0	94.0
5	2	93.0	96.0
5	3	78.0	75.5
9	1	56.5	74.0
9	2	74.0	75.0
9	3	44.0	54.5
9	4	81.5	84.0

well distributed cases than Dr. Starch used in his formation of the scale. The sentences from only two steps are given but several others show similar conditions, namely, that one or two of the sentences are much more difficult than the others within the same step.

5. The scale covers a number of types of grammatical errors but the author has furnished no data showing the basis of selection of the types and the relative importance of each.

Charters' Diagnostic Language Tests, one each for pronouns, verbs, and miscellaneous, attack the same problem attempted by Starch, but with a few noteworthy variations.

1. The material was secured from an extensive survey of language usage among children and was grouped on the basis of the three classifications noted above.

2. Instead of the place of error being pointed out and a choice made necessary between a correct and an incorrect form, the child is confronted with a series of sentences each of which he is told to mark with a cross in case he thinks it correct and to mark out the incorrect form and write the correct one in case a correction is to be made.

3. The test is scored in terms of single unit value for each sentence marked correctly.

4. The test for pronouns purports to present each of the different pronominal uses, and thus covers completely this phase of language ability. The following illustrations show how the test will point out clearly to the teachers the errors which the children fail to recognize as incorrect. The following sentences were pronounced correct by more than 75 percent of a large group selected at random from eighth-grade classes. The sentences are numbered as they appear on Dr. Charters' Pronoun Test.

3. It teaches a person something you may use.
5. When one lives in town they hear noises.
13. Are those them?
28. That's her.

The following sentences were pronounced correct by more than 50 percent of the same group of eighth-grade pupils.

7. Who do you want?
14. It was only us.
21. He pushed John and I.
25. Annie called to you and I.
27. Was that him?

The use of the test with grammar-grade pupils shows clearly that a number of these type errors persist with a very large proportion of the children who are practically ready to leave the grades, and the individual test papers will point out definitely to the teacher the work which needs additional attention. It is probable that the sentences should be given a weighted score on the basis of a combination of frequency of social usage and the inaccuracy of eighth-grade pupils, in order that their relative importance may be more forcibly impressed upon those who take the test.

A second edition of Charters' tests provides additional spaces beside the sentences. In this space the child is to write the grammatical reason for the changes which he makes in correcting the sentence. This affords a measure of the child's ability to give the correct grammatical rule or principle which applies to the case in point. He must not only know the principle but also be able to recall it in its proper relationship.

Kirby has attacked the same problem as Starch and Charters. He presents correct and incorrect forms for the child to choose between in the same manner that Starch does. Kirby, however, scores on the basis of unit value for each sentence. No evaluation

has been made for either difficulty or social usage. Here, as was stated concerning both Starch's scale and Charters' test, the children tend to reveal quite accurately the types for which they do not distinguish correct from incorrect language forms. In this test, however, no provision has been made to care for the operation of chance.

In addition to the check on the language form, Kirby attempts to measure the knowledge of grammatical principles involved in a different manner from that used by Charters. Beside a set of sentences there are given the different grammatical principles involved in this set of sentences. These, however, are arranged in such a way that a principle is never directly beside the sentence whose usage it governs. The child is therefore called upon not to recall from his own memory the grammatical principle which will justify the correction he has made, but to select from a number of grammatical principles which are before him the one which will apply in the particular case.

As was shown in the case of the Charters' tests some of the sentences involved language usage which is incorrectly checked in a large proportion of cases, while others are seldom missed. For example, it was found with two hundred seventh- and eighth-grade children that the sentence "It is (I) (me)" was missed but by 2 percent of the children; "(Mr. Smith) (Mr. Smith he) went ahead" by only 1 percent; "There (was) (were) many reasons for his actions" and "He (don't) (doesn't) belong in that group" by but 7 percent in each case. On the other hand, "He is the man (who) (whom) you said was injured," "I (lay) (laid) on the sand two hours yesterday," "It is a slight to me who (has) (have) always been your friend," and "We admire (that) (those) kind of people" were missed by more than 50 percent of the children on the language side alone.

On the technical grammar side, the principles governing the following sentences—"That book is not (hern) (hers), "(May) (can) I bring the next story to read?," "(Leave) (Let) me go with you," and "He divided his money (among) (between) his four brothers"—are the only ones which were correctly checked by practically all of the children. At present it would seem that the tendency in language work in our elementary schools is away from technical grammar with corresponding emphasis upon formation of correct language habits. Wherever this attitude is believed to

be the important one, the second phase of this test may well be omitted, since any test tends to emphasize the type of work which will get the results measured by the test. If the supervisor does not wish his teachers to stress the knowledge of grammatical principles, he should not measure the children's lack of knowledge of such principles.

It is true that ability to recognize incorrect language forms and to correct them does not guarantee that the child will use the correct form in either his oral or written language. Nevertheless it is quite certain that whenever the child cannot or does not recognize incorrect forms when presented to him for discriminating effort, he is not likely to use the correct form when his focus of attention is upon the general thought as is the case in either oral or written composition.

The purpose of this paper has been accomplished if the following facts have been clearly pointed out.

First, that the trend in the measurement of language has been from a gross measure of general merit to a specific measure of individual factors which are included in this general merit.

Second, that different efforts have been made to analyze the field and to discover through a study of social usage the phases which need careful attention.

Third, that tests which bring sharply to the attention of teacher and pupils the strength and weakness of each individual in the phase of language measured are more valuable than those which fail to reveal these situations.

Fourth, that while a considerable amount of work has been done, it is evident that our instruments of measurement are not yet perfect, that the field has not been completely covered, and that, therefore, although we are on our way, there remains yet a tremendous amount to be done.

INTELLIGENCE TESTS IN CLASSIFYING CHILDREN IN THE ELEMENTARY SCHOOL

CHARLES FORDYCE
University of Nebraska

For the purpose of determining to what degree the school supervisor may through mental tests predict the educational career of public school children and thus throw light on their proper classification, the writer made a study of the results of the Haggerty Intelligence Examination in comparison with the school grades and estimates of teachers in the case of a group of pupils in the elementary grades at Lincoln, Nebraska. The tests were given to 1,078 grade children distributed as follows: primary grades (I, II, and III) 602 pupils, intermediate grade (VI) 476 pupils. The Haggerty Intelligence Examination, Delta I was used in the primary grades and Delta II in the intermediate grade. The tests were given by Miss Clara Slade, psychologist in the Lincoln schools, and the results were tabulated and interpreted by a class of 67 graduate students in educational measurements under the direction of the writer. The test paper of each pupil was corrected by three different students working independently. The examination was given in the middle of May, 1920.

All teachers having to do with the instruction of the children gave their ratings of the intelligence of the pupils at the time the tests were made. The judgment of the teachers was based upon class grades made during the year and upon their general estimate of intelligence. In the case of each pupil the teacher's rating represents the average of the combined estimates of several instructors. As will be seen in Table I, the intelligence rank, as given by the teacher, was arranged on a five-point scale: very superior pupils being ranked A, superiors B, average pupils C, inferiors D, and very inferiors E. The rating by the intelligence tests was arranged on a similar scale and was based on the I. Q.'s obtained in the usual way (that is by dividing the mental age by the chronological age). Rank 1 in intelligence represents the students whose I. Q. was 120 or above, rank 2 those with an I. Q. from 110 to 119, rank 3 those with an I. Q. from 90 to 109, rank 4 those with an I. Q. from 80 to 89, and rank 5 those with an I. Q. below 80.

In Table I all entries in the upper left-hand corners of the compartments are for the sixth-grade pupils; those in the lower

TABLE I. CORRELATION BETWEEN TEACHERS' ESTIMATES AND INTELLIGENCE QUOTIENTS

I. Q.'s	Teachers' Estimates					Totals
	E	D	C	B	A	
	9 / 32 / 23	49 / 84 / 35	28 / 50 / 22	4 / 9 / 5	0 / 0 / 0	90 / 175 / 85
80						
	3 / 11 / 8	26 / 79 / 53	42 / 72 / 30	12 / 21 / 9	1 / 2 / 1	84 / 185 / 101
90						
	2 / 15 / 13	32 / 118 / 86	139 / 263 / 124	52 / 132 / 80	9 / 13 / 4	234 / 541 / 307
110						
	0 / 1 / 1	0 / 8 / 8	14 / 26 / 12	16 / 47 / 31	12 / 19 / 7	42 / 101 / 59
120						
	0 / 1 / 1	1 / 5 / 4	5 / 22 / 17	10 / 24 / 14	10 / 24 / 14	26 / 76 / 50
Totals	14 / 60 / 46	108 / 294 / 186	228 / 433 / 205	94 / 233 / 139	32 / 58 / 26	476 / 1078 / 602

Correlation, +0.44; P. E., 0.02

Figures in upper left-hand corners of squares are for sixth-grade pupils, total 476.

Figures in lower right-hand corners of squares are for pupils of grades I, II, and III-A, total 602.

Figures in center of squares are for all pupils, total 1,078.

right-hand corners are for the primary-grade children; while those in the center of the compartments are for all children—i.e. the entry in the center of each compartment is the sum of the other two entries in the same compartment. The detailed tabulation given in Table I shows that for 445 of the pupils or 41 percent, the ratings of the teachers and those of the intelligence examination were the same. According to the intelligence rating the teacher overestimated 383 cases of those of average or inferior capacity. This tendency is found to be common and due mainly without doubt to the fact that many of the average or inferior pupils are over aged. On this account, while they are inferior in mentality and may be doing only average work in the grade in which they are placed, yet they have the maturity of body and the instincts and emotions of children of their own age. They are therefore

estimated to a considerable degree by these characteristics. The overestimation is also probably due in part to the fact that inferior pupils are in many instances attractive, vivacious, and talkative, always ready to respond to the teacher. These qualities usually elicit a good grade.

Of the average or superior pupils 220 were underestimated by the teachers. This may be due partly to the fact that the superior student is in many instances young for the grade in which he is placed. He is, therefore, immature in bodily development and in his instinctive and emotional reactions, and is in some degree judged by these qualities. The underestimation of capable children may also be partly due to the fact that the work is so easy for the superior boy or girl that it does not challenge his efforts. As a result of this he often neglects his assignments, and is consequently rated average or low by the teacher.

It is a significant fact that among the 1,078 pupils, 74 percent of the highest ranks given by the teachers were given to pupils who were young for their respective grades. Such pupils had higher Intelligence Quotients, and being mentally superior to their fellows should be expected to do superior work. Ten percent of the highest grades were given to the over-aged whose intelligence quotients were low. This doubtless was due to the prevalent tendency on the part of teachers to overestimate the older pupils. Of those ranked average by the teachers 60 percent were of average rank as shown by the mental test, thus corroborating the fact usually found that the average pupil is more correctly judged and ranked by the teacher than either the superior or the inferior. Of the average grades given by the teachers 28 percent were given to pupils of inferior mentality as shown by the mental test. Of the lowest grades given, 4 percent were given to superior children. This is due no doubt to the tendency to underestimate the superior or younger pupils or to the fact that these superior pupils find the work so easy that they do not put forth an effort to do it. One of the greatest values of the mental test is shown in this latter case. It discovers the pupil who should be doing superior work and leads the teacher in many instances to bring about a procedure which will stimulate such students to do better work.

The highest rankings by the teachers were made in the case of those of highest intelligence quotients; of the students with

intelligence quotients below 90, 57 percent were given low ranking by the teachers, and only 9 percent received above average ranking. One-half of one percent were placed by the teachers in the highest rank.

The median I. Q. for the entire group was 100.5. The boys ranked higher in the mental tests than the girls, but the girls were rated slightly higher by the teachers. The median I. Q. for the boys was 105, and for the girls 96, but the median rating by the teachers was 85 percent for the boys, and 86 percent for the girls. It is difficult to explain this discrepancy in the correlation. The correlation between the ratings by the mental test and by the teachers was 0.44 with a probable error of 0.02. This correlation is not as high as we secured in the case of similar tests in the eighth grade and the high school.

Conclusion

This study in the elementary grades of the Lincoln schools indicated:

1. That the boys ranked slightly higher in the mental tests than the girls, but that the girls ranked somewhat higher by the teachers' estimates. (This is contrary to the results obtained in the Teachers College High School where the girls ranked superior, both by the mental tests and the estimates of the teachers.)

2. That the teachers tended to overestimate the inferior student and to underestimate the superior.

3. That the correlation between the mental tests and the teachers' estimates was sufficiently high to indicate that it is possible for one to discover by means of mental tests along with the grades and estimates of teachers such capacities as are essential in determining the proper grading and promotion of pupils.

THE INTELLIGENCE EXAMINATION FOR HIGH-SCHOOL FRESHMEN[1]

IRA J. BRIGHT
Superintendent of Schools, Leavenworth, Kansas

We had four general purposes in mind in carrying forward the study of the intelligence examination for high-school freshmen. The first was to determine the extent to which an intelligence examination can be used at the time a pupil enters high school, as a means of forecasting his probable success in the various high-school subjects and in order that he may be given intelligent guidance in his program-making. The second was to provide the best possible criteria on which to base the organization of class groups. The third was to provide the instructors in the high school with a definite statement of the mental abilities of the individual members of their respective classes, in order that they might better adjust their methods of teaching, the subject matter of their courses, and their requirements to the needs of the individuals in their respective classes. The fourth purpose was to determine the adaptability of the Terman Group Test of Mental Ability to first-year high-school pupils.

I am sure no argument is necessary among school men to prove the statement that up to the present time there has been very little guidance given to high-school freshmen in their program-making. On entering the high school the child has the opportunity to elect certain courses or subjects. When advice is given by advisory committees, class sponsors, or principals, it is given in most cases without any definite knowledge on the part of the advisor of the intellectual equipment of the child.

I know certain principals who always advise those who complete the eighth grade in their buildings to take Latin and algebra in the high school regardless of the mental capacity of the pupils. A review of the failures by subjects from the Leavenworth superintendent's reports for years past shows an excessively large number of failures in both Latin and algebra. This is no doubt due to the fact that a few years ago the distinctive disciplinary value of Latin and algebra was unquestioned. It was

[1] A paper presented at the meeting of the National Association of Directors of Educational Research at Atlantic City, N. J., March 3, 1921.

thought that these subjects were worth while for the general mental training which they provided regardless of all other considerations. The importance given them by tradition is still effective in drawing great numbers into these courses who are sure to meet with discouragement and failure because they lack the mental power to master them. That the loss in the freshmen year of high school, due to the lack of an intelligent and scientific method of guiding pupils in their program-making, is tremendous there can be no question.

The scholarship grades used in this study were those given by the high-school teachers at the end of the first quarter of this school year. The intelligence examination was not given until two weeks after the quarterly reports of the teachers were handed in and the results of the intelligence examination were not returned to the teachers until all information from the teachers prompted by this test had been secured. This precaution was taken so that there could be no conscious or unconscious prejudice influencing the teacher in the distribution of the scholarship grades or in answering questions concerning certain individuals about whom the writer sought further information.

The tests were all given by the high-school principal after he had gone over them with the writer and after he had made a careful study of the instructions. A stop watch was used so as to avoid discrepancies in the time element of the test.

The grading was all done in the superintendent's office under his immediate direction and supervision. The statistical work was done by the writer himself.

The correlation coefficients mentioned in this study have reference to the Pearson Product-Moment Coefficient of Correlation. Most of the coefficients were computed by the long method applying the formula $r = \frac{\Sigma\,(xy)}{\sqrt{\Sigma x^2\ \Sigma y^2}}$ and were checked by applying the shorter method suggested by Dr. Ayres in the April number of the JOURNAL OF EDUCATIONAL RESEARCH.

In the correlation tables showing the distribution of teachers' marks and intelligence scores, six steps were used for the scholarship grades and fourteen for the intelligence scores. In the correlation tables showing the distribution of scores in each of the ten tests of the Terman examination and in the complete test,

from ten to twelve steps were used for the individual tests and fourteen for the total score.

Table I shows the distribution of the freshmen according to the teachers' marks in Latin and according to intelligence scores. The median intelligence score of all Latin I students was 121.7 while the median intelligence score for all freshmen was 97. This shows that some selective agency had been operative. There were, altogether, nine in the freshmen class whose intelligence scores were between 40 and 49 on the Terman scale. Fortunately only one of these elected Latin. There were also nine freshmen in the next lowest group—i.e., those with intelligence scores between 50 and 59. None of them elected Latin. Out of the ten in the next lower group only two were taking Latin. At the other end of the distribution the result of selection is quite as noticeable. In the entire freshmen class there were three whose intelligence scores were between 170 and 179. All three of them were taking Latin. There were also three in the 160 to 169 group, and two of the three were taking Latin. In substance the students with the higher intelligence scores elect Latin while those with the lower intelligence scores do not. The selection here shown is due primarily to two causes. First the general feeling among pupils that Latin is hard. The brighter pupils not having had trouble in the elementary schools in making their grades are less effected by this feeling. Those of lower intelligence have more reason to be afraid of a hard subject. The other cause is the fact that we have given some attention the past few years to the guidance of pupils in the selection of their courses. We have attempted to keep out of the Latin classes those who are rated very poor by their principals. While the attempt to guide pupils in this matter has been inadequate, it has been somewhat effective.

The distribution on the intelligence scale of those making the different grades in Latin indicates in a general way that the correlation between teachers' marks in Latin and the intelligence scores is high. The coefficient of correlation is +0.65.

The complete distribution shows that 86 percent of all those whose scholarship grades in Latin were above 90, had intelligence scores above the 75-percentile, that is they were in the upper quarter of the whole freshmen distribution on the Terman Group Test. All first-year Latin students whose intelligence scores were below the 25-percentile received failing grades in that subject.

TABLE I. TERMAN GROUP TEST OF MENTAL ABILITY AND TEACHERS' MARKS IN LATIN

Intelligence Score	Teachers' Marks in Latin I					Total
	70–74	75–79	80–84	85–89	90–94	
170–179				1	2	3
160–169				2		2
150–159	1		1	2	1	5
140–149			1	1		2
130–139		3	1	1	1	6
120–129	1		7	2	2	12
110–119	2				1	3
100–109		2	1	1		4
90– 99	1			1		2
80– 89	2	3	1	2		8
70– 79	2	3		1		6
60– 69	1		1			2
50– 59						
40– 49	1					1
Total	11	11	13	14	7	56
Median	87.5	86.3	125	130	135	121.7

Only 15 percent of those whose intelligence scores were below the median of the group received scholarship grades in Latin above 80. Or stating it conversely, 85 percent of the first-year Latin group whose intelligence scores were below the median of the whole freshmen group received the lowest passing grade or failed altogether.

I believe therefore it is a safe prediction to say that those whose intelligence scores in the Terman Group Test are below 76, the 25-percentile of the entire freshmen distribution, have absolutely no chance to make a passing grade in Latin; and that those whose intelligence scores are below 97, the median of the whole group, will do unsatisfactory work in Latin. And I believe the prediction will hold good for almost any high school, that all freshmen whose intelligence rating according to this test places them in the lower quarter of the entire freshmen group will most likely fail, and that those whose intelligence scores are below the median of the freshman group will most likely do unsatisfactory work.

Let us now consider the distribution of English grades and intelligence scores as shown in Table II.

TABLE II. TERMAN GROUP TEST AND TEACHERS' MARKS IN ENGLISH

Intelligence Scores	Teachers' Ratings in English						Total
	70–74	75–79	80–84	85–89	90–94	95–100	
170–179					2	2	4
160–169						2	2
150–159				1	4	2	7
140–149	1	2			3	3	9
130–139		2	3	3	5	1	14
120–129			1	6	5	2	14
110–119	2	3	1	3	1	1	11
100–109	3	2	4	3	6	1	19
90– 99	3	2	5	7	2		19
80– 89	2	5	2	3	5	2	19
70– 79	2	7	7	6	1		23
60– 69	3	2	2	1			8
50– 59	1	1	2	3	2		9
40– 49	3	1	1	1			6
Total	20	27	28	37	36	16	164
Median	85	85	90	92	122	143	99

Table II in a general way indicates that there is some positive relation between teachers' marks in English and intelligence scores. The coëfficient of correlation is +0.72.

A more detailed study of the complete distribution will reveal some facts not precisely indicated in the Table II. It shows that 70 percent of all those who received the highest scholarship grades in English also had intelligence scores in the upper quarter of the entire freshmen distribution. On the basis of Table II the chance is about 1 to 25 that those whose intelligence rating places them below the upper quartile will receive the highest (95-100) scholarship grades in English. Moreover, of those receiving failing grades in English, 40 percent were in the lower quarter of the intelligence distribution and 70 percent were below the median.

These facts warrant the statement, I think, that the kind of grades those who enter the English classes of the first year of high school will likely make, can be rather well determined in advance by the Terman Test.

The correlation table (Table III) showing the distribution of algebra grades and intelligence scores tells much the same story except that the correlation is a little lower, the coefficient being +0.50.

TABLE III. TERMAN GROUP TEST AND TEACHERS' MARKS IN ALGEBRA

Intelligence Scores	Teachers' Ratings in Algebra						Total
	70–74	75–79	80–84	85–89	90–94	95–100	
170–179				1		2	3
160–169				2		1	3
150–159					1	5	6
140–149		2	2	1		3	8
130–139	1	2	3	2	4		12
120–129		1	1	2	4	5	13
110–119	3			1	2	2	8
100–109	2	3		4	4	4	17
90– 99	2	3	2	4	5		16
80– 89	4	4	5	3	1	2	19
70– 79	2	8	3	3	3	1	20
60– 69	3	4	1	1			9
50– 59		5	1	2			8
40– 49	2	1	1	1			5
Total	19	33	19	27	24	25	147
Median	86	78	87	99	107	127	98

From the more detailed distribution we find that 88 percent of those receiving the highest scholarship grades had intelligence scores above the 75-percentile of the entire intelligence distribution. Also, 71 percent of those who failed or received the lowest passing grade in algebra, had intelligence scores below the median of the whole freshmen group. Only 30 percent of those in the algebra classes whose intelligence scores were below the median of the whole freshmen group received grades in algebra above 85, that is, grades in the three upper scholarship groups.

It is obvious that a large number of our first-year high-school children are enrolled in algebra who, on account of their low intelligence, can never receive material benefit from it. There is practically no chance for those in the lower quarter of the intelligence distribution to do satisfactory work in algebra, and very little likelihood that those whose intelligence scores are below the median will get much value out of the subject. To my mind it is positively absurd to assume that all high-school pupils should take the course in algebra. Certainly if all are required to take this course, the subject matter, requirements, and methods should be adapted to the varying mental abilities represented in the freshmen group.

The distributions in the other academic subject are similar to those already discussed, and so I shall pass on to the distribution of grades in the handicraft subjects. I have included in this group manual training, domestic art, domestic science, freehand drawing, and penmanship. The coefficient of correlation between ntelligence scores and teachers' marks in handicrafts is only +0.36. This is very much lower than the correlations between the academic grades and intelligence scores. The significance of this fact cannot be escaped. It tells us plainly what to do with those of low intelligence who enter the high school. It is clear that this group should be encouraged to take handicraft courses and that their enrollment in Latin and algebra especially should be discouraged.

The high mortality rate in the first year of high school is no doubt due largely to discouragement which this group is bound to receive, especially in the academic subjects, under the present methods of administering the courses. When teachers can be provided w th an accurate index of the mental abilities of their pupils and when they have the ability themselves to adapt subject matter for the benefit of their respective classes, the number of failures will be greatly reduced if not entirely eliminated. We have as far as possible organized the classes for first-year high-school pupils who enrolled the second semester on the basis of intelligence scores. The enrollment in the midyear was small but we were able, for the most part, to put those whose intelligence scores were above the median of the group into one section and those whose intelligence scores were below the median of the group into another section. From all reports so far received the differ-

ence in the quality and quantity of work being done by the two sections is marked.

Where it is possible the classes in any subject should be organized on the basis of three standards; those in the upper quarter of the intelligence distribution comprising one division, those in the middle 50 percent of the distribution comprising another division, and those in the lower quarter, a third. The possibilities of the classes comprised of children of superior intelligence are almost unlimited. The individuals in these classes can be held responsible for a much higher quality of work as well as for a much greater quantity of work.

For the classes in the lower division the courses should be simplified and these groups held only to standards adapted to them. It is unreasonable to expect the best results in classes where there is represented such wide variation in mental abilities as is now found.

To my mind one of the greatest values which come from the use of intelligence tests is the stimulative effect of the result on the individuals tested, especially those making the higher scores. The recognition on the part of the individual that he stands high on the intellectual scale makes an appeal to his pride that is more effective than any other appeal that can be made.

Since giving the intelligence examination in our freshmen classes, the principal of the high school has called into conference those whose intelligence rating was high but whose scholarship grades were low, pointing out to them the fact that they are not in any sense measuring up to their possibilities and appealing to their sense of pride to raise the standard of their school work. This is proving wonderfully effective and will always prove effective when the principal or teacher is a person of strong personality and has the confidence of the student body.

On the other hand, I believe the greatest possible caution should be exercised in the use of the results of mental tests. In the first place, teachers must learn that the results secured from any one test may, in individual cases, be in error. No teacher should jump at the conclusion that because a child makes a low score in a test he belongs on the low intellectual plane indicated by the score. Judgment in these cases should be held in abeyance until the test result is verified or disproved by other tests and the further reactions of the individual. High scores in an intelligence

test are more likely than low scores to be true indications of mental ability. Then, too, an over-evaluation of mental ability is not nearly as serious as an under-evaluation of ability.

To get a further check on the validity of our test results we sent out to the teachers concerned a list of the names of the pupils who had high intelligence scores and low scholarship grades and of those who had low intelligence scores and high scholarship grades. The list was accompanied by a questionnaire in which the teachers were asked to indicate which of the following factors were most responsible for the scholarship grade received by the pupil: mental ability, attendance, health, indifference, laziness, interest, effort, attention, energy, affability, application. As an example let us take the report on one of these cases. Here is a boy whose intelligence score is 54 but whose scholarship grades are medium. Three of the teachers say that he hasn't much mental ability; one says he has ability. The first three teachers say that this scholarship grades are due to interest, application, and attention. In this case we have the judgment of three teachers against one that the test evaluation is about right.

To summarize the reports of the teachers, 85 percent of all those whose intelligence scores were in the upper quarter of the whole freshmen distribution but whose scholarship grades were low, were reported by the teachers in this "follow up" questionnaire as having mental ability. The low grades were accounted for in this report by lack of application, laziness, inattention, poor attendance, and the like. There were 29 in this group with high intelligence scores and low scholarship grades. Of these 25 were credited with high mentality in the "follow up" report. There was, then, a material disagreement between the test evaluation and the teachers' judgment in only four cases of the 45 in the upper quarter.

There were only seven of the 45 in the lower quarter of the intelligence distribution whose scholarship grades were above 85, that is, in the three higher scholarship groups. In the individual reports of the teachers, four of these were rated as having low intelligence. So there were only three of the 45 on which the intelligence test evaluation and the evaluation of the teachers differed.

In short, the Terman Group Test of Mental Ability has given us a distribution that is most consistent with the teachers'

judgment of mental ability as represented in the scholarship grades and as represented in the analysis of the individual cases.

It is evident, then, that the most influential factor in the determination of scholarship grades is mental ability. The greatest help that can be given a teacher is an accurate evaluation of the mental abilities of her pupils. If the classroom work of a pupil does not measure up to his mental ability, the cause should be immediately sought and eliminated. It is a tremendous loss to the individual capable of making the highest scholarship grade to be permitted to go through school with only average scholarship attainment. There is a tremendous loss to society as well as to the individual in every case where superior ability is not recognized and is not given the best possible opportunity for development.

I am thoroughly convinced that the efficiency of high-school teachers can be immeasurably increased by the judicious use of some such intelligence examination as the one used in this study. It is possible, however, that a shorter examination may be sufficient for the purpose. To determine whether or not the Terman Group Test could be shortened without impairing the results, we found the coefficient of correlation between each of the ten tests and the whole examination and between the first

TABLE IV. CORRELATION BETWEEN EACH TEST OF THE TERMAN GROUP TEST AND THE WHOLE EXAMINATION

Test	Correlation Coefficient
1	0.77
2	0.62
3	0.82
4	0.71
5	0.66
6	0.51
7	0.75
8	0.51
9	0.52
10	0.75
1–5	0.90

five combined and the whole. These correlations are shown in Table IV in the form of coefficients. Table V shows the details of the correlation between Tests 1 to 5 and the entire test.

TABLE V. TERMAN GROUP TESTS. CORRELATION BETWEEN TESTS 1–5 AND TOTAL SCORE

Scores on the Entire Test	Scores on Tests 1–5											Total
	16–23	24–31	32–39	40–47	48–55	56–63	64–71	72–79	80–87	88–95	96–103	
170–179										3	2	5
160–169									1	2		3
150–159								1	6			7
140–149							1	5	4			10
130–139						1	5	6	2			14
120–129					1	4	7	2				14
110–119				1	2	5	3					11
100–109				3	5	10	2					20
90– 99				4	9	8						21
80– 89			1	12	10							23
70– 79		4	12	8								24
60– 69		5	4	1								10
50– 59	5	4										9
40– 49	5	4										0
Total	10	17	17	29	27	28	18	14	13	5	2	180

These tables make clear, I am sure, that the first five tests may be used as an intelligence examination and that the results

will be nearly as dependable as they would be if the entire test were given.

CONCLUSION

1. That the probable success of first-year high-school pupils in the various courses can be predicted with a reasonably high degree of accuracy at the time they enter high school by the use of intelligence tests.

2. That the intelligence test clarifies the teachers' problems.

3. That the intelligence test offers the best criterion for the organization of class groups.

4. That the intelligence test affords guidance to teachers in the distribution of scholarship grades.

5. That the intelligence test, because of its stimulative effect on the individual, is a powerful agent of motivation.

6. That the results of intelligence tests must be used judiciously or otherwise great harm may be done.

7. That the Terman Group Test is well adapted to high-school freshmen.

8. That the last five tests of the Terman examination may be omitted without materially affecting the result.

9. In general, that the application of intelligence tests to first-year high-school classes is practicable and that unless we make use of them in improving the methods of teaching, in adapting the courses of study to the needs of the various groups, in guiding pupils in their program-making, in stimulating the individual to a greater realization of his possibilities, and in arousing in him a desire to measure up to those possibilities, we shall fall far short of our opportunities in these things.

Editorials

METRON

In July, 1920, a new international review called *Metron* was launched in Italy under the editorship of Doctor Corrado Gini, Professor of Statistics in the University of Padua. The second number appeared last December. This periodical is to appear as a quarterly and will amount to seven or eight hundred pages per year. It is published by the Industrie Grafiche Italiane at Rovigo (Veneto), Italy at fifty lire. At the rate of exchange at this writing this amounts to $2.04. This is indeed the time to buy foreign books.

As a literary venture *Metron* is unique. It accepts original articles in Italian, French, German, or English and prints them in the language in which they are written. Moreover, it is not exclusively or even particularly interested in a special field of knowledge. It is interested rather in the method by which authors investigate and report on problems, no matter what field they may represent. The one requirement is that papers shall represent a statistical treatment. Articles are also accepted on abstract methods in statistics. It is in this sense only that *Metron* may be said to have a specialty.

It is quite evident that the development of statistical inquiry has proceeded to the point where it is being erected into a field of knowledge if not into an independent science. This development has been going on for some time, and the publication of *Metron* is by no means the only evidence of it. For example, in 1913 Doctor Hugo Forscher wrote an important book on the theory of statistics which he entitled, "Statistical Method as an Independent Science."[1]

Originally, as the name implies, statistics had something to do with the state. It was not necessarily quantitative in character. As questions of state came to be more and more closely considered by an ever-increasing number of people, the necessity for a method of writing and reporting in a fixed language forced the adoption of numerical terms. Thus a rude procedure

[1] Forscher, Hugo. *Die statische Methode als selbständige Wissenschaft.* Leipzig: Veit and Company, 1913. 365 pp.

of a kind which we should now call statistical was built up. The application of this method in other fields naturally came about as the need for a rigorous treatment of their phenomena was felt. Thus we have social, economic, vital, and actuarial statistics, and later biological, psychological, educational, and financial statistics. There is now scarcely a field of knowledge in which statistics is not employed.

Of course, the field in which we are most interested is education. In it statistical methods have developed with a rapidity which has quite outstripped the ability of workers in education to comprehend it. Yet the time is coming when one can no more be a competent student of education without a working knowledge of statistics than one can be a competent actuary or economist without such knowledge. It is therefore altogether appropriate that a science of statistics should be recognized. It is not nearly so important that it be called a science as it is that it be accorded a definite position and appreciated as useful for its own sake.

The first number of *Metron* is a strong one. The first article (in French) is on "Statistical Method." It is a brief but well-organized general article on the subject. Important contributions are made by such distinguished statistical writers as Gini, Czuber, and Edgeworth. As indicating the wide range of topics, we may note the fact that one article (in English) belongs to the field of entomology and treats of the length of time which bees are away from the hive on each expedition, while another article (in French) deals with finance, and a third (in Italian) deals with horse-racing. This last article by Professor Gini is an interesting comparison between the success of book-makers at Rome and Milan as indicated by the relation between the actual order of finish of the horses and the order as expected according to the betting. The Roman members of the fraternity were more successful in picking the winners.

Our personal interest in a publication such as *Metron* has perhaps induced us to give it more space in these pages than our readers will consider appropriate. If we felt justified in devoting still more space to it, we might proceed to demolish our imaginary opponent; in other words, having set up our man of straw, we might proceed to knock him down. We feel quite sure that *Metron* is important as a statistical event. As such it should interest workers in all fields of knowledge to which statistics

apply. We are also sure that because of its polyglot character this publication is important for the graduate student in American universities because of the fine training it will enable him to obtain in the reading of foreign languages. We hope *Metron* will justify the fondest hopes of its founders. We admire their enterprise, and we wish them a full measure of success.

B. R. B.

THE LAW OF THE SINGLE VARIABLE

In the construction of educational tests there has been too little formulation of fundamental principles and too much mechanical application of statistical procedure. When, therefore, the author of a test accompanies the account of its derivation with a careful statement of the fundamental principles upon which she believes tests should be constructed, these principles deserve the thoughtful consideration of those interested in the construction and use of educational tests. Mrs. May Ayres Burgess in a monograph entitled *Measurement of Silent Reading* has formulated certain principles which she considers fundamental. These are epitomized in the title, "The Law of the Single Variable."

The three variables of a pupil's ability, or more strictly speaking, of his performance, are: (1) quality of performance; (2) level of difficulty on which it is given; (3) amount of work done within a fixed time, or rate of work. These variables are recognized as being complex but they are considered to include all factors which affect a pupil' performance.

Performances which depend upon a single variable are illustrated by the author in the field of athletics. In the high jump the variable is the heighth of the bar cleared. It represents the difficulty of the performance. The quality and the rate are either negligible or constant. In the races the rate is the variable, both difficulty and quality are either constant or eliminated. In shooting at a mark, quality is the only variable considered. In all of these cases the ability of a contestant is measured in terms of a single variable. The variable which is used is determined by the nature of the ability.

In applying this law of the single variable in the field of educational measurements the author states that it "consists of distinguishing the possible controlling, varying factors; devising means of holding them all constant save one; and measuring that

one." Correspondence with the author reveals that this statement was not intended to mean what it appears to mean. It appears that the law of the single variable was not intended to be interpreted as meaning that two of the variables must always be kept constant for all pupils while the third was measured. It should be interpreted to mean that pupil performances depend upon three variables, or have three dimensions. In describing a performance these three variables must be recognized. If a variable is not constant for a group of pupils its variation must be recorded for each child in the process of testing and when scores are interpreted they must be on the basis of a single variable. For example, if children are to be compared with reference to their rate of work it must be shown that both the quality and the difficulty of the work done were the same for all pupils. "What the law of the single variable does not permit is the attempt to compare combinations of the three variables in unknown and varying amounts."

It is obvious that the law of the single variable should not be interpreted to mean that when a test is given all pupils should be forced to give a performance which is constant with respect to two of the three variables. The characteristics of the abilities which children acquire are not restricted to a single variable. When a group of pupils is performing in a given subject-matter field in the way which is most natural for each, large individual differences are exhibited with respect to these three variables, particularly with respect to rate and quality of work. This is true even in the case of groups of pupils which have received the same instruction. It is reasonably easy to control the difficulty through the selection of the content of the test. The control of the rate of work, or the quality of work is not so simple. In certain fields it may be shown that one of these factors has little effect upon the performance. For example, in spelling the rate of work has little effect upon a pupil's performance, provided the time allowed is sufficient for all pupils to write the words. In some fields the rate is an unimportant variable and can be neglected. This would be true in the case of painting or drawing when the products are real works of art. In these cases the ability of the performer depends upon the single variable of quality.

If a variable is controlled by an arbitrary procedure which produces unnatural conditions it appears likely that the ability

which functions is modified, or factors are introduced into the testing procedure that produce the same result upon the performance. For example, if in the case of handwriting all pupils of a group were forced to write at a fixed rate, those who were accustomed to write more rapidly or more slowly than this rate would exhibit abilities different from those which normally function in their handwriting. To the extent that this is done the performance secured would not be a true index of the ability of the pupil

Although the author fails to make it clear in her monograph the "moràl" of the law of the single variable appears to be that measuring instruments must explicitly recognize the possible existence of the three variables, or dimensions, of a pupil's performance. These must be described separately if accurate interpretation of the scores is to be possible. If a variable is omitted in a pupil's score it must be shown to have been constant in its effect upon the performances of different pupils or to be socially unimportant. This is a principle which has been sadly neglected by many of the makers of educational tests. In addition to yielding measurements which are likely to be erroneously interpreted, instruments in whose construction this principle is violated also lead to implications that are not in agreement with our educational objectives. Grade norms for a test are generally interpreted as standards or goals to be obtained in the respective grades. An instrument which measures the pupil's ability in terms of the most difficult step reached therefore implies that our objectives in that field are to prepare the pupil to do more and more difficult things. We should teach pupils to do not merely difficult things but things that are socially useful. Instruments which fail to measure the rate of work in fields where it is an important variable are also open to criticism. They do not tell the whole truth about a pupil's ability.

Dr. Burgess has rendered a real service in formulating the principles embodied in the law of the single variable. It is unfortunate that they are not stated in such a way that misinterpretation would not be likely. Whether or not future makers of tests are prepared to accept the law of the single variable it is hoped that they will feel its incumbent upon them to formulate carefully the principles of measurement upon which the construction of their instrument is based. W. S. M.

Reviews and Abstracts

E. H. Cameron, *Editor*

Burgess, May Ayres. *Measurement of silent reading.* New York: Department of Education, Russell Sage Foundation, 1920. 163 pp.

In this monograph the author describes the derivation of a new type of scale for measuring ability in silent reading, which is called "Picture Supplement Scale (P. S.—1)." Supplementary to this account the author summarizes certain general considerations relative to the theory of educational measurements. The Picture Supplement Scale is the product of a series of carefully planned experiments. The chapters devoted to the theory of educational measurement are also evidence that much thought was given to the fundamental principles on which the construction of educational measuring instruments should be based. In both cases the author has made a notable contribution to our literature in this field.

The wide-spread interest in educational measurements has resulted in the derivation of an exceedingly large number of measuring instruments within the past few years. For the most part the authors of tests have accepted without critical consideration, methods of test construction which were originally developed in a single field and have applied them without any critical analysis to the field in which they happen to be interested. In the midst of such work it is refreshing to read the account of the derivation of the Picture Supplement Scale. Statistical tables and the description of an elaborate statistical procedure are conspicuous by absence. The major portion of the monograph is devoted to a critical consideration of the principles which should govern the construction of tests with particular reference in the field of silent reading. The central theme of the part devoted to the theory of educational measurements is summarized in the caption "The Law of the Single Variable." The writer comments upon this law editorially elsewhere in this issue of the Journal of Educational Research and for that reason only brief reference will be made to it here.

The function of the "Picture Supplement Scale" is to measure the amount of printed material which the pupil can read within a given time. In respect to the type of reading which is measured the author claims that it is "a test of careful reading." The scale consists of twenty exercises equal in difficulty and printed on one side of a 12 x 19 sheet. Each exercise consists of a short paragraph and a picture. The paragraph is partly descriptive of the picture and in part consists of directions for drawing a supplementary picture. The test of the pupil's reading of the paragraph is the drawing of the supplementary picture. The pupils are allowed five minutes to do as many as they can. Directions for scoring are liberal. Any sort of a drawing which shows that the pupil followed instruction is to be counted as correct. The scale is recommended for use in grades III to VIII.

In the derivation of this scale the author identified twenty-five factors which influence a pupil's performance in silent reading. An attempt was made either to eliminate or to control twenty-four of these factors, leaving only the amount read as the variable factor. The ability to draw the pictures required (difficulty of action

demanded) is announced as being held constant but one naturally questions its constancy. Some of the drawings to be made are very simple. For example, in Exercise 6 a pupil is asked to cross out a portion of a picture. In other exercises the drawing required is more complex. In Exercise 4 the pupil is to draw a picture of three feathers. No instructions are given to the pupil concerning the quality of the drawings required One would, therefore, expect to find that some pupils would attempt careful drawings while others would make very hasty sketches.

The instructions for administering the tests are meager and do not specify the exact explanations to be given to the pupils. The pupils are given no preliminary exercises to acquaint them with the nature of the tests. One would, therefore, expect the administration of the tests to lack objectivity. One is also inclined to question the objectivity of the scoring, even though the directions state that any sort of a drawing which shows that a pupil has followed instructions is to be accepted as correct. In fact one state bureau which is distributing the Picture Supplement Scale has found it necessary to issue detailed directions for scoring the exercises.

The pupil's score is the number of exercises done correctly in the five minutes allowed for the test. The author considers that both the difficulty of the exercises and the quality of reading required are kept constant. The uniformity of difficulty is secured by use of exercises for which the same percent of correct responses is in general obtained. The quality of the reading is considered to be kept constant because the pupil is given credit for only those exercises which he answers correctly. Hence, according to the author the score is a measure of the amount of reading which the pupil has done in five minutes or, in other words, his rate of reading.

It is likely that the number of exercises done correctly furnishes the best single numerical description of a pupil's performance on this test. However, this score is not a measure of his rate of work. It is a combination of rate of reading and quality of reading. This combination is not the same for all pupils. For example, a pupil may make a score of 10 doing ten exercises with 100 percent accuracy. Another pupil may make a score of 10 by doing twenty exercises with 50 percent accuracy. The rate of reading is not the same for these two pupils. Neither have they read with the same quality.

The pupil's score in terms of the number of exercises done correctly is to be translated into a score on a percent scale. This scale is different for the different grades. For example, if a pupil is in the fourth grade ten paragraphs done correctly entitle him to a score of 68. If he is in the seventh grade his score is 50. No real advantage is gained by this translation and it will doubtless be confusing to many.

The reliability of the scale was studied by having it given twice to a few small groups of pupils. The reliability of coefficients calculated for these groups closely approximate those for a number of other reading tests.

The Picture Supplement Scale represents an ingenious and serious attempt to construct an instrument for the measurement of silent reading ability. It deserves a place among our best reading tests. If the present form were supplemented by more detailed directions for administering it and by the addition of a few preliminary exercises, it would be well adapted for general use, particularly in the lower grades.

W. S. M.

CUBBERLEY, ELLWOOD P. *The history of education.* New York: Houghton Mifflin Company, 1920. 849 pp.

CUBBERLEY, ELLWOOD P. *Readings in the history of education.* New York: Houghton Mifflin Company, 1920. 684 pp.

Professor Cubberley has, through the exceedingly useful character of his educational work, secured a very large following who will welcome the publication of these two important contributions to the history of education. Professor Cubberley is fortunate in having attained to an almost unique position in the educational world. Although for many years he has been directly connected with university work he has kept in the closest touch with the developments of educational administration and, both through his writings and his work in various cities as an educational surveyor, he has personally contributed powerfully to the general acceptance by administrators of standards in educational administration. Having retained this intimate and constructive attitude toward education over so long a period of years one expects to find his treatment of the history of education modified by this point of view.

This expectation is realized, as he himself points out in his preface. It is difficult for one who is continually surveying the work of public school systems not to see the very close connection which always exists between a civilization and its type of schools. The development of educational theory, as such, is forced to the background while the consideration of the actual efforts in education is correspondingly emphasized. It is probably due to the rather general realization of the indefiniteness in the relation between educational theory and the solution of present educational problems which has caused during the last few years the sharp decline in popularity of the traditional courses in the history of education. At any rate, we have in these two books the concrete efforts of a man who is primarily interested in the effort of modern society toward realizing itself effectively through improvement in school organization, to prepare a history which would treat education as a phase of the general development of civilization. That he has accomplished this in a way which will win the hearty approval of the school men of the United States almost goes without saying.

Were one inclined to be over-critical, he might question the wisdom of including within the pages of the text so much that is found in any well-written survey of history. He might ask himself whether the book could not easily have been made much smaller and yet not be reduced to any degree as a presentation of the development of education. I found myself continually asking this question; and yet, when I specifically attempted to cut out topics that primarily belonged to general history I hesitated, realizing that in spite of the greatest efforts of teachers of history, much of this information might be so poorly mastered by the student that the real significance of the educational topics would be lost without this other presentation. So I have finally withdrawn my criticism believing that from the teaching standpoint Mr. Cubberley justifies the inclusion of this material.

All of Mr. Cubberley's texts are unusually readable. These volumes are no exception to the rule. I predict that many schoolmen will, from the motive of pleasure alone, read the text from cover to cover.

The *Readings in the History of Education* have been carefully selected and are so arranged as to follow the chapter arrangement of the text. In this way much that might otherwise have had to be included in the presentation is immediately available in the second volume. So well has this collection of readings been compiled that none

but a specialist in the history of education would feel much need for access to sources not found within the covers of this book.

Professor Cubberley follows the customary organization of material as found in texts now generally used but never loses sight of the point of view stated in the subtitle of the volume "Educational practise and progress considered as a phase in the development and spread of western civilization." He divides his study of western civilization into four parts: "The Ancient World," "The Mediaeval World," "The Transition from Mediaeval to Modern Attitudes," and "Modern Times." He, however, appreciates fully the vastly greater significance to the student of education of the modern period and devotes approximately half the book to part 4. Students of history will doubtless question the propriety of dismissing the educational contribution of Greece, Rome and early Christianity with a paltry hundred pages. Possibly one might query whether complete omission might not be justified with better grace than the cursory mention which, frequently, is all that is included. I, personally, however believe that the distribution of emphasis is well justified. Unless prerequisites are insisted upon, complete omission produces with many readers distorted perspective, while cursory mention suggests the influence upon modern conditions which is all for which the author is striving.

Certainly he is justified in assuming that the student or reader is not completely ignorant of ancient and mediaeval history. Therefore he is equally justified in assuming that in this volume the chief purpose is to show the vital connections between the significant development of the civilization of Greece and Rome and the educational conditions of those ages, and that in turn Christianity not only furnished the only institution which could serve as a foundation for modern civilization but conserved certain fundamentals both of civilization and education without which the task of reconstruction would have been far more complex and difficult. Cubberley's treatment of the significance to modern education of the revival of learning, of the Protestant revolts, of the development of scientific inquiry and of the scientific method, while conforming to that usually found is clear and convincing and dramatically real to the average reader.

School superintendents and principals will be especially impressed with the discussions centering around the abolition of privilege, the rise of democracy, the new theory of education, and the struggle as a result of which the state takes over the school.

Those who have read his *Public Education in the United States* will find much that is duplication of the earlier work. This is probably inevitable. Professor Cubberley could not afford to omit important and pertinent topics merely because he had already discussed them in another volume. At the same time, those who read all of his writings may well be pardoned if they regret that he did not see his way clear to vary more his presentation of these topics. The same feeling has occurred to me more than once in reading *Public Education in the United States* and comparing it with his *Public School Administration*. Topics in school administration although legitimately and necessarily found both in a history of education that professes to include the educational movements that have had their best exemplification in the United States and in a special treatment of education in the United States surely can be developed from somewhat different angles if they frankly make their appeal to the same group of readers. However, this is really a very minor criticism. Most of us will profit by rereading these topics and those who use only one of the volumes receive the point of view in each case that the author is earnestly desiring to impart.

Again the scientific historian might question the propriety of including so much that is so recent that it is still in the process of making. I, personally, however, find in this one of the chief charms of the book. Cubberley makes our history of education a real pulsating thing. He shows how the older movements affect present conditions and how just as in the past our educational organization is a method through which our people become better adjusted to the conditions under which they live and which they must in turn affect.

I commend these two volumes to teachers of the history of education in our departments of education in universities and to teachers of the history of education in our normal schools as handbooks admirably designed to give a student of this subject a point of view towards education which is eminently sane and constructive. Too many teachers and even administrators fail to realize the cause of many of our present educational problems. Studies of this character develop the power to see the relationship between a given problem and the given social, political, and economic conditions. In my judgment, therefore, a course in the history of education from this point of view is worth while for all students of education to take and I would not be at all surprised to see as a result of the publication of these volumes a very real increase in the influence of the history of education as a vital division in our departments of education.

C. E. Chadsey

University of Illinois

Cubberley, E. P. *The history of education*. Boston: Houghton Mifflin Company, 1920. 849 pp.

In his preface, Professor Cubberley tells us that he has "not tried to prepare another history of educational theories," and that he has "omitted reference to many theorists and reformers." His dismissal of Plato and Aristotle in a few sentences, and without giving even the briefest sketch of their schemes of education, is equivalent to omission of reference. The reader is told that Quintilian was "the foremost Roman writer on educational practice," and that his *Institutes of Oratory* gives "a detailed explanation of the old Roman theory of education at its best," but is left to wonder what "the old Roman theory of education at its best" was like. "For such omission I have no apology to make." *Ipse dixit.* Very well! This review, then, to borrow the author's phraseology, will be "less concerned . . . . with the educational and philosophical theories advanced by thinkers . . . . than with what was actually done." However, it should be observed that a well-balanced History of Education, or History of Civilization, should contain liberal reference to "theories."

As one reads the text he is conscious of passing through many areas of condensation and rarefaction. The reviewer questions the advisability of subjecting the casual reader, or immature student, for whom the work is designed, to so many changes of pressure.

The preface, and the book itself, suggest that it may be more profitable to search for errors in the statement of "facts," than in the presentation and evaluation of "theories." And their "name is legion."

On page 7, in the context having reference to the opening of the "Christian era," appears the statement that "Greek was forgotten." The fact is, Greek was not forgotten at that time. The following page contains the information that "the study of Greek and Hebrew" was "revived" during the "Italian Revival," and on page 95, in a

section devoted to the "Rejection of pagan learning in the West," (fourth–sixth centuries) we read that "the Greek language was forgotten, and was not known again in the West for nearly a thousand years." This incorrect impression is enhanced by the statement (page 424) that "Greek . . . . was restored to the western world" during the period embraced by the years 1333–1433. The study of Greek was not unknown in Sicily and Italy in the twelfth century, and the "study of Greek and Hebrew" occupied a fairly prominent place in the thirteenth century. John of Capua, Raymond Lull, Siger de Brabant, Guillaume Bernardi di Gaillac, and Arnauld de Villaneuve knew Greek and Hebrew. The Dominicans made provision for instruction in these languages in the thirteenth century, and the Council of Vienne, 1312, decreed that the Eastern tongues be taught at Rome, Bologna, Salamanca, Paris, and Oxford. It is incorrect to say that "during the Middle Ages . . . . none could read it" (Greek). The "study of Greek" continued, without interruption, throughout the Middle Ages.

A legend that no longer obtains among those who have done intensive research in the field of mediaeval literature is continued in such expressions as the following: "the long intellectual night of the Middle Ages," "the long intellectual night of the mediaeval period," "the long period of barbarism and general ignorance," and "the long period of intellectual stagnation."

True enough, during the seventh and eighth centuries, "Many of the priests were woefully ignorant," but many have been "woefully ignorant" ever since that day. An influential few, however, were not so ignorant. And one must take *cum magno grano* the statement that "the Latin writings of the time contain many inaccuracies and corruptions which reveal the low standard of learning even among the better educated of the clerical class." The standard of the educated clergy was high; it was "low" only among the uneducated clergy. Would it be proper to conclude that a "low standard of learning" obtains among present-day writers, of college texts even, because their "writings . . . . contain many inaccuracies and corruptions"? Certain canons of historical method should be observed in our evaluations and interpretations.

The conclusion (p. 263) that "Mediaeval education . . . . prepared for the world to come; not for the world men live in here" is too sweeping. Only the narrowly religious education of the period "prepared for the world to come." Secular education, whether for the nobles and the well-to-do, or for those of "low degree," did prepare "for the world men live in here." "The world of the mediaeval monk and the Scholastic" (p. 279) was not the whole world of the Middle Ages, by any means.

At one point (p. 155) we read that theology was "the one professional study of the whole middle-age period," and at others (pp. 199, 211) that medicine and law were "professional subjects" also.

For the convenience of the reader of this review five consecutive sentences, that occur under the title "Results of their work" (the work of the Schoolmen, p. 192), are here reproduced. "The work of the Schoolmen was to organize and present in systematic and dogmatic form the teachings of the Church. This they did exceedingly well, and the result was a thoroughgoing organization of Theology as a teaching subject. They did little to extend knowledge, and nothing at all to apply it to the problems of nature and man. Their work was abstract and philosophical instead, dealing wholly with theological questions. The purpose was to lay down principles, and to offer a training in analysis, comparison, classification, and deduction which would prepare learned and subtle defenders of the faith of the Church." Evidently the author wishes to give the impression that the Schoolmen dealt "wholly with theological questions."

What could be further from the truth! Scholasticism was interested in every phase of intellectual activity. Only part of its work was that of "reorganizing and systematizing theology."

A rather striking area of condensation contains the following, in two sentences: "Luther's . . . . translation (New Testament, 1522) . . . . virtually fixed the character of the German language. Calvin's *Institutes of Christianity* (French edition, 1541) . . . . fixed the character of the French language, and Tyndale's translation of the New Testament (1526) . . . . fixed the character of the English tongue." Sweeping, to say the least! And these statements occur without qualification.

On one page (258) we find that Magellan "circumnavigated the globe" in 1519–21, and on another (424) that his "ships rounded the world" in 1515–18. A similar discrepancy occurs in connection with the dates of publication of Rousseau's *Social Contract*, and *Émile*. "In 1752 appeared both his *Social Contract*, and *Émile*" (483). Page 508 indicates that they were published in 1762. In another instance (p. 525) the author says that "In 1799 . . . . Jefferson tried unsuccessfully to secure the passage of a comprehensive bill . . . . for the organization of a complete system of public education for Virginia." If this is a typographical error for 1779, it should have been corrected in a footnote to page 526, and in the *Readings*, page 427. Again, an incorrect impression might be gained from the statement (p. 524) that "New York in 1787 created an administrative organization known as the University of the State of New York." But, with the author's *Public Education in the United States* at hand, the reader is fairly safe; for here (p. 258) he learns that the University of the State of New York was "established in 1784, and organized in its permanent form in 1787."

The author quotes, without comment, on page 369, the following, from Draper: "All the English schools in the province (of New York) from 1700 down to the time of the Declaration of Independence were maintained by a great religious society . . . . called the Society for the Propagation of the Gospel in Foreign Parts." As a matter of fact, but very few of the schools in New York during this period were maintained by the S. P. G. And it is equally incorrect to say (p. 521) that "In the Middle Colonies, where the parochial-school conception of education was the prevailing type, the school remained under church control until after the foundation of our national government." Most of the schools of the Middle Colonies, in the eighteenth century, were private-venture schools, and not "under church control." Again, a faulty implication is contained in the assertion (p. 520) that "By the close of the colonial period the new American Academy (p. 463), with its more practical studies, had begun to supersede the old Latin grammar school." Schools offering these "more practical studies" far outnumbered those of the Latin Grammar School type long before "the close of the colonial period."

The average instructor, as well as the average student, accepts the printed word. If care is observed, to point out certain errors, the text is usable. Its scheme of organization is good. It supplements admirably the work that has been done by Graves and Monroe.

Robert F. Seybolt

University of Illinois

News Items and Communications

This department will contain news items regarding research workers and their activities. It will also serve as a clearing house for more formal communications on similar topics, preferably of not more than five hundred words. These communications will be printed over the signatures of the authors. Address all correspondence concerning this department to Walter S. Monroe, University of Illinois, Urbana, Illinois.

Educational Research in Oklahoma

A letter from Henry D. Rinsland, Director of Department of Research and Guidance, Ardmore (Oklahoma) City Schools, gives the following information concerning educational research activities in his state.

At the invitation of County Superintendent Mrs. M. O'Daniel Rinsland, a survey is being made of the public schools of Johnston County. This is the first county survey in Oklahoma. The members of the survey staff are from the University of Oklahoma and two of the state normal schools. The survey has been organized in seven divisions as follows:

Division 1. Organization and Administration
Division 2. Buildings and Equipment
Division 3. Attendance and Enrollment.
Division 4. Instruction, Course of Study
Division 5. Teacher Status
Division 6. Tests and Measurements
Division 7. Finance.

Mr. Rinsland also reports the extensive use of intelligence tests in the public schools of Ardmore under his supervision. He has also spent some time in Henrietta supervising the giving of educational and intelligence tests in that city.

Bureau of School Service University of Kansas

At the University of Kansas, Dean F. J. Kelly of the School of Education conceived the idea of a bureau whose function would be primarily service rather than research. At the present this Bureau of School Service is under the direction of F. P. O'Brien. A recent letter from Professor O'Brien gives the following information concerning the work of this bureau.

"The Bureau has directed a survey of the curriculum and building problems of the Lawrence school system. The report was completed early in January and will soon appear in printed form. The Bureau is at the present time cooperating with the schools in cities of the first, second, and third classes, in making a study of the teacher-salary situation in the schools of the State. The report is nearing completion and will soon be distributed to the school systems in this State so that it will be in their hands before the time for the re-employment of teachers for the coming year. The Bureau is also conducting a survey of the school situation as it is found in several of the school districts in an adjoining county of Kansas. Some of these districts are in a condition of distress due partly to the fact that several small high schools are maintained where one should be adequate. The financial burden of the present arrangement is crushing for certain districts. They made an appeal thru the County Superintendent to this

Bureau of the University for some professional assistance in advising them what to do. That whole section of the county is now anxiously awaiting the report of our investigation.

"It is evident that such undertakings are chiefly of the service type but the Bureau has also been promoting Research Work in a smaller way by assisting and encouraging students interested in experimental or research study."

Latin Tests by a Latin Specialist

Dr. Evan T. Sage, Professor of Latin of the University of Pittsburgh, is developing some Latin tests. In a recent letter he says, "We are going on this assumption for the time being that in addition to certain qualities which we may not be able to measure a student needs some knowledge of forms, vocabulary and syntax. There are of course tests in existence to measure all these things separately. But as the concrete aim of Latin teaching is ability to translate it should be possible to make a translation scale which will measure achievement in all these things at once and also serve as a basis for investigation of the other elements involved in successful translation. The tests of this sort which are now in existence seem to be open to the objection that they are not based on the standard vocabulary and standard syntax. Unfortunately inflections have not yet been standardized though we are working on that problem."

It is quite within the bounds of reason that a Latin test by a professor of Latin may take precedence over every such test now available. We have long thought in the testing of abilities in subjects taught in secondary and higher educational institutions the experts in these subjects should be heard from. It takes something more than a mere scale maker to devise acceptable instruments of this sort. We hope Professor Sage will carry his investigations to a successful conclusion.

A Bulletin on Silent Reading

We are in receipt of a "Bulletin on Silent Reading" from Superintendent F. S. Camp of Stamford, Connecticut. It is addressed to his principals and teachers. It is one of the best we have seen. Both silent and oral reading are divided into intensive and extensive activities, each of which is defined and illustrated.

"Silent intensive reading" is largely study. "Silent extensive reading" is reading for pleasure and individual interests. "Oral intensive reading" includes in the early grades the conventional mechanics of reading and in the upper grades the study of masterpieces and "group study reading," i. e., reading in which all children have the same texts before them and are actuated by a common purpose. "Oral extensive reading" occurs when one person reads to a group of listeners.

The bulletin gives for each of these four types of reading activity a suggested proportion of time to be utilized in each of the eight elementary grades. The shift of emphasis from grade to grade, so far as the program may show it, is thus clearly set forth in the table on page 70. The entries indicate the number of tenths of the entire reading time in each grade.

Superintendent Camp discusses the question of reading from these four points of view and provides definite means by which the principals and teachers may check up their success in meeting the objectives he sets up. Several blanks for recording the amount and character of the reading done by children are provided. This sort of a bulletin cannot fail, especially when followed up by the supervisory staff, to produce *measurable* improvement.

SUGGESTED TIME VALUES, TEN REPRESENTING THE BASE FOR READING

Type of Reading Activity	Grades I	II	III	IV	V	VI	VII	VIII
Silent intensive reading	2	1	[a]	[a]	[a]	0	0	0
Silent extensive reading	0	1	2	3	5	6	7	7
Oral intensive reading	7	6	5	4	3	2	2	2
Oral extensive reading	1	2	3	3	2	2	1	1

[a] Large time values here in grades where the child must *learn* from the printed text; but as it is not considered as falling under "reading exercises" in the ordinary sense, no ratio numbers are assigned.

Schoolmen's Week University of Pennsylvania

Schoolmen's Week at the University of Pennsylvania was held Thursday, Friday, and Saturday, April 7–9. There were 1,885 persons registered while 8,565 attendances were noted in 31 programs. The number of registrations and attendances were both larger than in any preceding year, the previous highest records, 1,392 and 3,802 respectively, having been made in 1920.

Among the speakers from outside the state were, J. H. Kirkland, Chancellor, Vanderbilt University, Nashville, Tennessee; Frank Aydelotte, President-elect of Swarthmore College, now Professor of English, Massachusetts Institute of Technology; John W. Withers, former Superintendent of Schools of St. Louis, now Dean of School of Education, New York University; Walter S. Dearborn and John M. Brewer, Harvard University; William S. Gray, School of Education, University of Chicago; Professor N. L. Engelhardt, Otis W. Caldwell, and Miss Fannie Dunn, Columbia University; and C. J. Galpin, Specialist in Farm Life Studies, Department of Agriculture, Washington, D. C.

Social and recreational features took the form of visitation of the campus and buildings, including the museum and library. The guests of the university were given luncheons on Thursday, Friday, and Saturday at noon, and on the evenings of Thursday and Friday. On Saturday over 650 luncheons were served, a reception was tendered by the Acting Provost, Doctor Josiah H. Penniman after the Friday evening program. Several hundred accepted the invitation of the Athletic Association to witness the baseball game between Pennsylvania and Swarthmore, Saturday afternoon. On Thursday night two hundred superintendents and supervising principals were given stage seats at the Academy of Music and the Metropolitan Opera House where the American Legion gave a program of unusual interest for the purpose of stimulating the spirit of Americanism.

It was generally considered the most successful meeting yet held. This result was due in large part to the faithful and cordial cooperation of the members of the General Committee, the Reception and Registration Committees, and the Advisory Committeemen, and to the support given by State Superintendent Thomas E. Finegan and City Superintendent Edwin E. Broome and their assistants. The executive officers

were the same as in previous years—Harlan Updegraff, Chairman, General Committee; Arthur J. Jones in charge of Secondary School Conferences and LeRoy A. King, Secretary, General Committee.

Equivalence of Forms I and II of the Burgess Picture Supplement Scale for Measuring Silent Reading Ability

Soon after the beginning of the present semester, the Burgess Picture Supplement Scale for Measuring Silent Reading Ability, Form I, was applied to about 1,000 seventh- and eighth-grade pupils with the following results:

Class	No. Tested	Paragraphs (Median)	Credit (Median)
VII-B	238	7	35
VII-A	263	8	38
VIII-B	242	9	41
VIII-A	219	9	38

It was expected that later in the semester Form II of the same scale would be applied for the purpose of securing a measure of the progress that is being made under present methods of instruction.

In order to get an idea of the practice effect of taking the test once, a group of high sixth-grade pupils was given Form I on April seventh. Under exactly the same conditions so far as it is possible to secure such a situation the same group was given Form II on the following day. The records of all pupils not present at both tests were discarded.

As in the case of the seventh and eighth grades, this group fell below standard, securing a median score of only 43. But to our surprise the median on the following day with Form II jumped to 64. Forty individuals made gains totaling 198 paragraphs, four neither lost nor gained and two lost a total of 2 paragraphs.

These gains were so notable that it seemed as if they could not be attributed entirely to practice effect. Therefore another group of 45 high sixth-grade pupils was selected and given Form II, followed on the second day thereafter by Form I. Instead of an improvement appearing as the result of experience there was manifested a decided loss. The median score dropped from 55 to 49. Twelve pupils gained twenty paragraphs. Nine neither gained nor lost and twenty lost a total of 55 paragraphs. This helped to cónform our suspicions that the gains of group one were partly due to a difference in difficulty of the two forms and not altogether to practice effect.

To further satisfy ourselves as to whether or not Form II is as difficult as Form I, a third group of high sixth-grade pupils was given Form II and after a brief rest, Form I. The general results agreed with the facts already noted. This class scored 58 on Form II and after a few minutes rest fell to a score of 49 on Form I. Seven pupils made a total gain of 10 paragraphs, over the record made on Form II, seven pupils made the same score and eighteen pupils lost a total of 43 paragraphs.

Our brief experience with this test may be summarized thus: (1) Form I appears to be considerably more difficult than Form II; (2) inasmuch as all of our Form I

median scores are below standard and all of our Form II median scores are above standard we may be justified in a suspicion that Form I standards may be a trifle high and that Form II standards may be too low. Of course, we claim for our results no values beyond that of the proverbial straw that indicates which way the wind is probably blowing.

Information concerning the experiences of others bearing on the main questions here involved will be eagerly awaited. Meanwhile we shall be in doubt as to how to complete our undertaking to secure a correct measure of the semester gains made by our seventh- and eighth-grade classes.

H. C. Daley

Survey Department,
Highland Park Public Schools,
Highland Park, Michigan

The Illinois Examination

While there is much discussion both pro and con in regard to tests students of education, supervisors, and teachers may be interested in reading of the use of the Illinois Examination in Bureau County and the Princeton city schools in October, 1920. The writer has been actively engaged in working with this battery of educational tests in Bureau County since the sixteenth of last August. The examination in this county involved three hundred eighty-five teachers and about five thousand five hundred boys and girls. It is impossible to give any account of the Princeton city schools without discussing the work of the entire county.

County Superintendent George O. Smith of Bureau County is not only a firm believer that educational tests will be a big factor in the solution of future school problems but he is also anxious that his teachers be prepared to participate actively in the use of them. As a result the writer offered work on educational tests two periods a day in the August institute. The Illinois Examination and Teachers' Handbook were made the basis of instruction. Two sections of about one hundred teachers each were formed. The work of each section was similar. Consultation periods were also provided. The work was given under the following divisions:

1. Each teacher was given the test in exactly the same manner in which she was expected to give it.
2. Each teacher was taught how to score the papers using her own as a sample.
3. The interpretation of mental age and intelligence quotient in terms of ability to do work of the grade in which the child was located was explained.
4. The interpretation of achievement ages and achievement quotients was considered with reference to planning suitable remedies for unsatisfactory instruction.

After giving tests, teachers were asked to make out record sheets like the one below. These are on file for reference at any time pupil promotions are being discussed. The column marked "Health" is to be marked for each pupil with 1, 2, or 3. One stands for excellent health, two good health, and three poor health. The health statement is filled out by the school nurses. The remainder of the sheet is filled in for each child from his examination.

Many criticize tests because we do not follow up. In the case of Bureau County and the Princeton city schools the follow-up work was carefully planned and carried out. The entire county was divided into districts for the convenience of the teachers.

BUREAU COUNTY PUBLIC SCHOOLS

ILLINOIS EXAMINATIONS GIVEN October 7, 1920

…ool Lincoln (Princeton) *Teacher* T.V.Fisher *District* 115

Names of Pupils (alphabetically)	Health	Grade	Chronological Age yrs. mos.	Mental or Achievement Ages					Intelligence or Achievement Quotient				
				Intelligence	Arithmetic	Reading			Intelligence	Arithmetic	Reading		
						Average	Comprehension	Rate			Average	Comprehension	Rate
…thur Rhoek	3	5	14	10	7	10	8	12	75	71	100	80	120

A called teachers' meeting was held in each district. At the meeting the test was carefully reviewed and the typical type errors called to the attention of the teachers. First of all, teachers were urged to secure a textbook[1] that would deal with the subject of remedies in detail. The arithmetical errors we arranged as to type. It was found that in arithmetic pupils were generally weak in the fundamental operations. To improve this condition drill cards[2] in arithmetic were secured. The teacher was then instructed to go over the test papers carefully and let each child work on those drills in which he did not make satisfactory scores. Each teacher is also provided with a sheet of problems showing the different abilities required in the fundamental operations. This same idea was carried out in all phases of the work in arithmetic. It will be readily seen that in this way the pupil received individual attention and was not required to drill or work on material that he could handle successfully.

The grade medians for the Princeton city schools are approximately equivalent to the grade standards in the case of general intelligence. This means that the boys and girls in the different grades of the Princeton city schools have, on the average, about the same capacity to learn as that exhibited by the average corresponding grade in other places. This does not mean that every boy or girl is up to standard with respect to general intelligence. We have some pupils who differ from it by as much as three years. In grade VII we have mental ages ranging from nine to eighteen years. The least range is found in the third grade.

The median scores for silent reading were up to the standards. This verifies several other tests that have been given in the last two years to the pupils of the Princeton city schools and is probably due in a measure to the methods used in the teaching of the subject.

[1] Monroe, W. S. *Measuring the results of teaching*. New York: Houghton Mifflin Co., 1918.
[2] Clark, M. G. Arithmetic Habituated, Series A.

The median scores in arithmetic are not satisfactory to the teachers nor are they up to the standards of the examination. There are several known causes back of this:

1. The test involved, in certain grades, operations that are not introduced in the Princeton city schools until later in the life of the child.

2. Previous to the present school year the arithmetic text was not satisfactory nor practical.

3. The instruction was not suitable as it was based entirely on the text in use.

The first of these causes cannot be overcome in our course of study. The second has been remedied by a change in arithmetic textbooks. The third is being worked out as rapidly as possible. At present some improvement has been shown. The superintendent has gone over the pupils' papers with the classroom teacher and the errors have been checked. Drill is being given each day on the basis of the individual needs.

The use of the Illinois Examination in Bureau County and the Princeton city schools has suggested the following observations with reference to its value.

The Illinois Examination is simple enough in its operation to make it a valuable aid to both rural and city schools. It is very easy to administer.

Teachers who have used it are inclined to diagnose more carefully all cases of failure to do good school work.

Teachers, upon discovering that this work is constructive and not for the purpose of finding fault with teachers, cooperate very willingly.

While the examination is standardized for the group and was not used for promotional purposes it has designated certain children as worthy of larger opportunities to make progress.

A better understanding of school situations is developed between teachers and supervisors.

The examination forms a preliminary basis for the study of gradation or grouping by mental rather than chronological ages.

It will be no small factor in making teachers realize that there is a scientific side to teaching that cannot be neglected.

Carleton Bloss Smith

Superintendent, Princeton, Illinois

National Association of Directors of Educational Research

(E. J. Ashbaugh, *Secretary and Editor*)

DOES EDUCATION PAY FOR ITSELF?

Recently at an educational meeting, a university professor spent forty-five minutes in elaborating the proposition that the public should be taxed to support only those forms of education which clearly return full economic value to society and that the individual should pay in full for those other forms of education which do not make such returns. He upheld the public elementary schools in general though insisting that their course should be changed in such a manner that they might well be called schools of citizenship. He suggested that perhaps these schools should carry the youth to about the age of sixteen by which time the principles of government and economic law would be so well grounded that the children so taught would hence forth be good citizens.

He indorsed public support for the graduate schools with emphasis upon the research laboratories, holding that they had already demonstrated that they return full value for the investment made by public taxation. Concerning the two remaining phases of education which are now largely supported by public taxation, secondary and undergraduate higher, he had serious doubts of the propriety of continued support. He expressed these doubts in such a manner that his hearers recognized that he had no doubts in his own mind. He was quite sure that the public could not continue the present trend of publicly supported educational opportunity for ever-increasing numbers, and that high-school and college education does not pay for itself in returns to society at large; perhaps not even in returns to the individual receiving it.

I am not concerned in this column with a debate of the subject. I do not conceive that this is the place or the time for either a rebuttal of his argument or a further elaboration of the correctness of his position. But I do believe that the thought expressed may legitimately be used as a basis of serious thought and comment apropos of the work of members of bureaus of educational research.

Let us face the question: not does education pay for itself but does my work pay for itself? Is the city or state for which I am working getting an adequate economic return for the expenditure which I am causing it to incur? Will this bit of investigation, that bit of research, these tabulations, those data gathered and interpreted yield a return in financial savings or increased efficiency in the system which can be made apparent to the school administration and the public in general? There are scores and scores of problems which have not been solved. There are many, many questions to which we must now answer that we do not know. There are all sorts of interesting things which may be found out about schools and education. But it is quite probable that in the present period of financial depression and restless questioning, the public—the paying public—is but little interested in pure research, in the answer to idle questions, or in the solution of problems which do not appeal to it and which you may not have demonstrated to be of real vital worth.

I do not believe that the public insists or will insist that the evaluation of all educational effort shall be expressed in economic terms. I am quite sure it is interested in spiritual values. But I am also quite sure that this same public is interested and will more or less definitely insist upon being shown returns for the investment which is made in terms which it can appreciate and which it will approve.

One of the members has written to your secretary for his opinion on the following points. Cannot the membership of the association contribute and thus enable him to answer with a composite judgment instead of an individual one? The questions are as follows:

1. From the standpoint of state administration, would it be advisable to have the same age definitions for intelligence test records and for reporting age-grade census data?

2. If one standard is fixed, is it better to consider an eight-year-old norm as including all the children of ages 8 years 0 months to 8 years 11 months inclusive or should we consider an eight-year-old as being anywhere between 7 years 7 months and 8 years 6 months inclusive?

3. In establishing age norms on either of these bases, is it better to group children in year groups or in half-year groups?

Dr. Franzen in charge of the research work in the Des Moines, Iowa schools sends in a twenty-point program of his work there. There is space for but a mere statement of the main points.

1. City-wide measurement of intelligence
2. City-wide measurement of arithmetic
3. Intensive measurement in three schools in terms of Educational Quotient
4. Measurement of product in high schools
5. Measurement of intelligence of incoming freshman class in one High School
6. City-wide measurement of handwriting
7. Formulation, etc. of test for incoming teachers
8. Construction of objective composition scale
9. Construction of scales, typewriting test
10. Construction of intelligence test for kindergarten children
11. An experiment in the ethics of school children
12. Measurement of oral and silent reading in grades I, II, and III
13. An experiment in the pedagogy of shorthand
14. Comparative statistics on athletic records
15. Course in statistics and measurements for principals and supervisors
16. Help in diagnosis of difficulties
17. Meeting with teachers whenever information or advice will make better teachers
18. Conferences with supervisors
19. Determination of a workable and reasonable salary schedule
20. Computation and arrangement of tables and of technic which will increase efficiency of office routine.

Dean Russell of the State University of Iowa has called a week's conference of superintendents who had not planned to attend summer school to meet at the

university about the middle of July to work intensively on problems of school finance. About a dozen men volunteered at the recent Conference on School Supervision to attend and work with him on this problem.

Several subcommittees of the standardization committee reported at the closed meeting on February 26, 1921. The report of the subcommittee on statistical methods was read by Dr. Rugg in the absence of the writer Dr. Truman L. Kelley. This report follows.

REPORT OF THE SUBCOMMITTEE ON STATISTICAL METHODS OF THE STANDARDIZATION COMMITTEE (TENTATIVE)

There is a variety of procedure in the treatment of test and measurement data. This is unfortunate from the standpoint of ready interpretation of published studies, but it is indicative of the virility of the movement. The number of workers each believing in the superior merit of his method of presentation makes it undesirable, without more information than is now available to your committee to take issue with any of the methods employed, except as implied in proposals (1) and (2). For the reasons stated and because statistical procedure in psychology and education should conform with that already established in the older biological and physical sciences your committee thinks it desirable to offer general rather than specific proposals:

1. Any worker presenting a new procedure should definitely recognize that it is impossible to prove the superiority of his method by reporting data interpreted by means of his method alone. Superiority is a relative matter and it is necessary to compare a method with alternative methods before its superiority can be established.

2. Any worker presenting a new procedure should definitely expect that the burden of proof lies with him. This implies that he should consider himself responsible for proving: (a) that his method is more reliable in the sense that the probable errors of results obtained by it are less than those obtained by the more standard methods. Proving this would give the method prior claim to excellence and it would then devolve upon other claimants to establish the superiority of their methods. Or, (b) that the method has advantages of interpretation or of expedition that more than compensate for its lessened reliability. This implies that the amount of this lessening in reliability is established and made known by the worker.

The following suggestions as to various measures are offered:

3. Of the measures of central tendency the arithmetic average or mean is easily understood, easily calculated, and usually for such distributions as are found in educational research work, has a smaller probable error than the median or mode. The median, as proved by educational studies of the last decade, is calculated in many ways. Its probable error is almost always greater than that of the mean, and has seldom been reported. Its brevity (not its accuracy) of calculation and ease of interpretation recommend it but in case these are not prime considerations your committee recommends the use of the mean in place of the median or mode as a measure of central tendency.

4. Of the measures of dispersion the standard deviation usually has, for such distributions as are found in educational studies, a smaller proportionate probable error than the average deviation or the quartile deviation. Its general use in scientific presentations as a measure of dispersion is recommended. For popular presentation

0.6745 times the standard deviation, or the probable error, is recommended. If necessary considerations have lead to the use of the median instead of the mean as a measure of central tendency, it is recommended that a measure of dispersion based upon percentiles, e. g. the quartile deviation, the 10–90 percentile range, etc., be used as the measure of dispersion.

5. Of the measures of correlation, the Galton-Pearson product moment coefficient of correlation is well established, having, in the case of rectilinear relationships, the smallest probable error of any of the coefficients of correlation. The Spearman ρ-coefficient based upon the squares of the differences in rank holds a similar position in the case of ranked data. These two coefficients are recommended because of their known properties and particularly because their probable errors are available. A measure of relationship for which a formula giving the probable error is not available should be considered as merely in the experimental stage.

Many devices have been proposed for the measurement of non-rectilinear relationships. No recommendation with reference to them is offered.

6. As a general statement of policy to be followed in educational research your committee would recommend that the most reliable of several available measures (such, for example, as measures of central tendency, of dispersion, of correlation, of frequency, etc.) be used, and that when one departs from this rule he should feel it incumbent upon himself to justify his procedure upon substantial grounds, such as the necessity for using a less time consuming method, greater simplicity of interpretation, etc.

TRUMAN L. KELLEY
Chairman, Subcommittee on Statistical Methods

At the closed meeting of the association at Atlantic City on February 26, 1921, Mr. Courtis, chairman of the standardization committee appointed at the Cleveland meeting, presented material gathered from the questionnaire which he had sent out to the members. On the following Thursday he made the report of the committee and the report was ordered printed. We present it herewith.

REPORT OF THE STANDARDIZATION COMMITTEE

Tabulation of the returns from the questionnaire sent to members revealed two outstanding facts: (1) a practically unanimous sentiment in favor of the publication of an official list of terms, procedures, etc. and (2) a fear that the attempted standardization may be premature. Therefore the statements following have been prepared as the recommendation of the committee, but it is suggested that the list be printed in the JOURNAL OF EDUCATIONAL RESEARCH as tentative standardizations for trial purposes only, final approval to be deferred until the annual meeting in 1922. During the interval use of the approved forms is to be experimental and voluntary, but members preparing material for publication are asked to try to conform to these regulations.

1. Members of the association preparing material for publication, and wishing to use alternative forms or definitions are asked to give reasons or explanations in footnotes.

2. A test composed of elements of uniform difficulty, or of several cycles of uniform difficulty, and used to determine the rate at which the work is done, should be called Rate Tests. (*Illustrations:* Cleveland Arithmetic Tests, Starch's Reading Tests, Courtis Writing Tests, etc.)

3. The term "scale" should be applied only to tests or material graded in difficulty, or quality, and used to measure degree of difficulty or quality.

4. A test composed of elements graded in difficulty and used to determine the most difficult test material the subject can handle successfully, under the prescribed conditions, should be called a performance scale. (*Illustrations:* Trabue's Completion Test, Language Scales, Gray's Oral Reading Scale, Thorndike's Reading Scale.)

5. Whenever the objective product of a pupil's test is measured as to quality by comparison with samples of known value, the collection of samples of known value should be called a Product Scale. (*Illustrations:* Thorndike Handwriting Scale, Hillegas Composition Scale, Rugg's Lettering Scale.)

6. In connection with both rate tests and performance scales the score obtained is frequently evaluated in terms of the results of giving the test to age or grade groups. This evaluation consists in placing the score upon a scale of difficulty. Where this form of scaling has been made possible, the testing instrument itself ~~is~~ often called a scale. (*Illustrations:* Burgess Reading Scale, Ayres Spelling Scale, Hahn-Lackey Geography Scale, etc.) At this time the committee is not able to recommend a suitable name for scales of this type, or for other tests and scales which are mixtures or combinations of the simpler forms named above.

7. The term "standard" should be applied to scores as an adjective, and only when the scores are set up as goals or objectives of teaching effort. If adopted by an authoritative organization, they may be called "standard scores" without qualification but scores set up as goals by an author should be designated as "author's standard scores." A score representing an actual performance of a group should be called a "norm." Whenever norms are published they should be accompanied by statements of the number of cases, upon which each norm is based and of the statistical procedure by which it was derived.

8. The term "average" should be used only in its generic sense to include any and all measures of central tendency. The term "arithmetic mean" should be used as the name of the sum of a series of measures divided by the number of measures. (*Illustration of correct use of term:* Averages in common use are the mean, the median, and the mode.)

9. The term "mean" when used alone should be taken to connote the "arithmetic" mean. All other types of means should be used with descriptive adjectives, as harmonic mean, geometric mean, etc.

10. There is need to distinguish between measures of central tendency derived from actual scores and those derived from frequency distributions. For instance, the new term "midscore" is suggested as a name to be applied to the middle most measure (or mean of the two middle measures when there are two), while the term "median" should be reserved for the value derived from a frequency distribution by interpolation. (See Rugg's *Statistical Methods Applied to Education*, pages 103-14.)

11. The term "Performance" should be used to connote the score obtained in a test. When, however, there is a desire to convey the idea of comparison with norms or standards the term "achievement" should be used. The two terms are, thus, not precisely synonymous.

12. The score of an individual in a test should not be considered as a measure of his ability but as a record of his performance under the particular conditions under which the test was given. Judgments as to ability are always inferences from performance and the greater the number of trials or tests from which ability is inferred, the greater the probable correctness of the inference.

13. The term "capacity" should be restricted so that it may connote just one thing: namely, inborn, or undeveloped possibilities of behavior.

14. Ability should be regarded as the product of experience acting on capacity.

15. The movement for the use of measurement in education has had for its goal the attainment of uniformity and certainty of interpretation. Great advances have already been made but still greater advances are possible. To aid in bringing about improvement, makers of tests should be careful to investigate the effect of every factor involved in standardization, conforming so far as possible to the law of the single variable. In general the process of standardization of a test will involve the following operations (although not necessarily in the order given):

A. Preparation and selection of test material
B. Experimental organization of test and instructions for giving the test.
C. Trial of tentative test to determine value of elements, gross validity, reliability, and optimum conditions of giving, scoring, etc.
D. Final organization of test.
E. Final formulation of conditions under which test is to be given, scored, tabulated, and interpreted.
F. Official determination of validity
G. Official determination of reliability
H. Official determination of norms.

The association accordingly invites makers of new tests to file with its standardization committee for publication data which throw light upon

A. What a test measures.
B. How reliably it measures.
C. The principles of formulae of construction.
D. Statements of conditions for giving, scoring, and tabulating the test and for interpreting the results.
E. Norms.

16. Members of the association and others are invited to send to the committee suggestions for the modification of the above, or other definitions or practices which should be standardized.

Problems

Two of the most important types of problems in measurement are those connected with the determination of what a test measures, and of how consistently it measures. The first should be called the problem of validity, the second, the problem of reliability.

Members are urged to devise and publish means of determining the relation between the scores made in a test and other measures of the same ability: in other words, to try to solve the problem of determining the validity of a test.

Members are also urged to study carefully the relations between the scores made in a test and the scores made in other trials of the same or equivalent editions of a test (the problem of reliability).

Norms should probably be set only after a careful study of the effects of variation of each of the factors of age, grade, time of year, heredity, maturity, training, and social status. As rapidly as proves feasible correction tables for variations in each of these factors should be worked out and published.

Respectfully submitted, B. R. Buckingham, W. A. McCall, A. S. Otis, H. O. Rugg, M. R. Trabue, S. A. Courtis, *Chairman*, *Special Standardization Committee.*

Journal of Educational Research

OFFICIAL ORGAN OF THE NATIONAL ASSOCIATION OF DIRECTORS OF EDUCATIONAL RESEARCH

VOL. IV SEPTEMBER, 1921 No. 2

CONTENTS

PUBLISHED FOR THE BUREAU OF EDUCATIONAL RESEARCH • COLLEGE OF EDUCATION • UNIVERSITY OF ILLINOIS •

Journal of Educational Research

Published for the Bureau of Educational Research, University of Illinois

THE JOURNAL OF EDUCATIONAL RESEARCH is published at 450 Ahnaip St., Menasha, Wis., monthly except July and August by Geo. Banta, Official Printer to the Public School Publishing Co. The subscription price is $4.00; $4.30 in Canada or Mexico; and $4.50 in other foreign countries.

The Public School Publishing Company conducts the subscription work. Send all orders to them at Bloomington, Illinois.

The JOURNAL OF EDUCATIONAL RESEARCH is both needed by and needs the subscriptions of many more teachers and school officers.

Articles to Appear Soon

A Cycle—Omnibus Intelligence Tests for College Students by *L. L. Thurstone.*
Analysis of Reading Ability by *S. A. Courtis.*
Measuring Progress by Means of Standardized Tests by *S. S. Brooks.*
Arithmetic Ability of Men in the Army and Children in the Public Schools by *Arthur Kolstad.*
Correlation: How to Work It on the Adding Machine by *Herbert A. Toops.*
Rate of Progress in Teacher Preparation by *W. Randolph Burgess.*
A Series of Standardized Diagnostic Tests for the Fundamentals of Algebra by *E. R. Douglas.*

JOURNAL *of* EDUCATIONAL RESEARCH

VOLUME IV SEPTEMBER, 1921 NUMBER 2

THE MEASUREMENT OF TEACHING EFFICIENCY[1]

J. B. SEARS
Leland Stanford Junior University

THE PROBLEM

The measurement movement.—During the past decade we have been busily engaged in the development of a new terminology for education—a quantitative terminology. We have been trying to measure the different machinery, processes, and products of the schools, and by these measurements not only to standardize, but also to rationalize every step in our procedure. Pedagogical and mental tests and scales, building scales, units of cost, hygienic standards, age-grade-progress norms, teacher rating systems, and standardized college entrance examinations, readily suggest to our minds the many aspects of education to which we have applied this quantitative language.

The whole measurement enterprise in education is very young, and so there can be no discouragement to thoughtful people, either in the fact that most of our standards are crude, and as yet only partially reliable, or in the further fact that much of the field is still unexplored. On the other hand there is substantial hope in the fact that the new methods and standards are being put to use in the schools as rapidly as they are worked out, and that everywhere the results of their use speak in unmistakable terms of their practical contribution to education.

Our subject here has to do with the measurement of teaching efficiency, and represents one of the many problems which together have made up the important movement which we have briefly suggested above. Judged, either by the demand for a purely objective scale of measure, or by the extent to which the "general impression" method of rating teachers still dominates in

[1] Read before a meeting of the department of secondary education of the Minnesota State Teachers Association at St. Paul, November 4, 1920.

practice, it is fair to say that this problem has not yet found a satisfactory solution. Judged, however, by the scientific work that has been done on the subject, and reported in the bibliography of 55 titles here appended, one is inclined to say that the solution is not far distant. Even a cursory reading of this literature shows clearly three things: first, that practical school men are persistent in their demand for such a device; second, that the methods for studying the subject have become increasingly scientific; and third, that the time is now ripe for a careful experimental study, which should result in the formulation of a rating scheme to be formally adopted and used by the teachers of the country.

Practical aspects of the problem.—Just what is this problem? From the standpoint of school practice it may be briefly stated as follows: In teaching as in everything else, the quality of work varies all the way from excellent to very poor, or even vicious. In our nearly three centuries of experience in education we have finally come to a fairly general acceptance of the principle that the school should reward excellent and penalize poor service. This can only be done satisfactorily when we find a means of stating in exact terms the degree of success attained in a given instance. We have tried in numerous cities to base promotion on merit and almost invariably the attempt has finally broken upon the rock of "what is merit?" General impression and mere opinion have done their best and failed, and we are able to recognize their failure. If we are to establish the principle that merit counts in our profession, then, what we need is a satisfactory means for measuring merit.

It is not from this angle alone, however, that we are beginning to recognize an insistent demand for a solution to this problem. The work of training teachers is becoming increasingly important, and as the process becomes more scientific the need for a means of measuring the growth of the teacher in training becomes urgent.

What a saving in energy would be effected, what financial waste would be checked, what an amount of justice would be established, and what a professional stimulus would result, if we had tests or instruments of measure by means of which we could predict the success of an applicant for teacher training work or for a teaching position, measure the rate of progress of the teacher in training, and evaluate the work of teachers in service.

An equally genuine need exists, too, for scientific instruments and methods that will aid in the diagnosis of the teaching and supervisory processes. It is not merely standards that we need, but, also, a more intimate and exact knowledge of the processes and products to be standardized. Supervisors of teachers in training, as well as supervisors of teachers at work, are beginning to recognize this need, and the reason they recognize it is because they are attempting to rationalize the weakest process in education, viz., that of supervision. To measure, then, in order that we may define and recognize merit; and to diagnose, in order that we may rationalize and perfect our processes of teaching and supervising, these are the educational needs which reveal to us the practical aspects of our problem.

Theoretical aspects of the problem.—On the technical or theoretical side the problem is that of defining "teaching success." This calls for an analysis of the teaching process, a process which we all recognize to be very complex. The teacher instructs, manages, and disciplines children, within the limitations of a certain environment over which she has at least partial control. There are certain general and many very specific educational objectives, in terms of which she must carry on her work. The amount of success is determined by the extent to which and the degree of economy with which right teaching objectives are attained.

Shall we try to define "success" in terms of the teacher's own qualities and virtues, or in terms of definable results which she produces in the classroom with children, or in terms of both? Any test by which we hope to predict the success of a teacher, or of a student teacher, must obviously deal with the personal and professional qualities of the teacher herself, since as yet there are no results to measure. And, since teachers are not generally in control of the same class for two years in succession it would seem practically necessary in all cases to have some measure of the teacher as well as of results.

What results shall we measure, and what of the qualities and traits of the teacher shall we measure? In what proportion should the two stand in our final success score? In the measurement of personal traits is our final score to be the sum of equal amounts of a number of traits, or is a complicated system of weighting necessary to express the true total value of all the

items involved? Add to these questions the necessity for a measure of the reliability of our final score once we have established it, and we have before us the technical aspect of our problem.

History of Teacher Rating Schemes

The need for more exact methods of defining "teaching efficiency" has been evidenced in the general literature on school administration for a very long time. But the discussion of ways and means for measuring teaching efficiency dates back only a very few years. The writer has found no study that bears at all directly upon the subject that was made before 1905, and no carefully devised rating scheme appeared before that of Elliott in 1912.[2] That does not mean that city superintendents had not been trying to base their appointments and promotions upon merit before this time nor that they had not attempted to analyze teaching success. They had been doing both of these for some years, but merit had been merely estimated in terms of general impressions with no attempt at scientific measurement.

Indirect studies of teaching success.—In 1905, W. F. Book, and in 1907, H. E. Kratz, sought to inquire into the elements of success among high-school teachers by making a study of the opinions of high-school pupils. This indirect method of approach was made from slightly different angles by Littler in 1914, who studied the failures of elementary-school teachers; by Moses in the same year, who studied the failure of high-school teachers; by Buellesfield in 1915, who studied causes of failures among teachers in cities of various sizes; by Anderson in 1917, who collected judgments on the relative importance of 15 different factors; and by Colvin in 1918, who studied the most common faults of beginning teachers in high school. The questionnaire method characterizes most of these studies, each of which was intended to throw some light on the factors essential to success in teaching. The statistical treatment of much of the data collected in these studies was good, and the results are of some value in that they tend to confirm our previous general impressions as to what are the weak points in the teaching process, and

[2] In order to save frequent repetition in footnotes and at the same time to bring together a fairly complete bibliography on this subject all such titles are listed at the close of the article.

to offer some suggestions to those engaged in training teachers; but they are negative in their approach, and at best they do not add greatly to our knowledge of the real factors in teaching success.

Studies of teaching success.—A second group of studies has approached the subject from the standpoint of success rather than failure. These studies deal with the judgments of school people, and differ from the above group not only in the point of attack but in the fact that they offer more thorough statistical treatment of the results. The first of these studies was made by Ruediger and Strayer in 1910, and was followed by those of Boyce in 1912 and 1915, Clapp in 1915, Anderson in 1917, Landsittel in 1917, Bradley and Moody in 1918, and Fordyce and Twiss in 1919.

In all these studies the attempt is made to show the relation of certain individual factors in success to general merit. Different statistical methods are used by different studies but all speak in terms of correlations.

The character of these contributions.—In some cases the judgments on which these studies are based were made in answer to a questionnaire, in others they had been recorded in the form of school grades which, in most cases, are little different from general judgments. An attempt has been made to analyze "general merit" or "teaching success." In doing this each writer has arbitrarily chosen such terms as he believed would express clearly recognizable qualities of the teacher, or clearly recognizable factors in teaching efficiency. In these terms there is variation, both as to number and name, as well as in the matter of organizing them into main and subordinate divisions. Some express their findings in terms of correlations only, others convert their correlations into scores after the fashion of Elliott's analytical score card.

If the results of these studies do not prove conclusively that teaching ability can be analyzed and expressed in objective terms they strongly suggest that it can be. They have attempted to find out whether we all mean the same thing when we refer to a teacher's "personality," "power to discipline," "power to manage," etc., and further, what is the relationship between each of these items and "general success." And the burden of the

evidence is strongly affirmative. The contribution of these studies is, therefore, a contribution in method and technic on the one side, and in actual analysis of teaching success on the other. And even if there is still some disagreement in results, and much overlapping in terms used; even if the devices proposed are still vague, unreliable, and too cumbersome for general use, surely this brief analysis of the work thus far done must convince anyone that a substantial foundation has been laid.

Other types of studies.—There are two other studies of somewhat different type that must be included in any review of the scientific work done on teacher measurement. These are the studies of Jones and Rugg. In 1917 Jones conceived the idea of trying to describe the results of a teacher's work by testing the mental behavior of her pupils. He chose two elementary classes in educational psychology which had been taught by two teachers, one of whom emphasized memory, and the other reason. Two tests, one useful in measuring memory, the other reason, were given to these classes. The experiment is too small and too inadequate in other ways to be at all conclusive, but the results are suggestive, since the points of view of these two teachers are clearly reflected in the results of the tests. Rugg has devised a measuring device patterned after the army rating scale, in which merit is expressed, not in absolute quantity but in terms of rank order. A teacher is rated on each item by comparison with a group of five other teachers who have been chosen as illustrating five degrees of efficiency ranging from very poor to very good. This plan of measurement has proved practical in the army and, within certain limitations, should work in education.

Characteristics of scales now available.—Of the rating scales now available for use some are merely off-hand analyses of general merit, while others represent a more careful analysis, together with an attempt at a defensible weighting of the different factors involved in general merit. Some have the score expressed numerically, others, by comparison, or by relative position on a scale of values. Of the latter, some provide for three possible ratings of each item (as poor, medium, good) while others recognize as many as ten divisions of merit. Some refer to the different items of merit by name, others by a question, and in either case

detailed definition of the item may or may not be given. Finally some are useful largely as devices for inspection and supervision, others especially as self-rating plans.

Our Next Step in Teacher Measurement

The problem of teacher measurement presents itself to us in two important aspects, the practical and the theoretical. In the one case we must clarify in our minds the functions that are to be served, the practical criteria to which the make-up of such a device must conform, and the how and where of its use. In the other we must work out an analysis of "teaching success"; so that we shall be able to define the elements essential to such success and we must be able to define the relationships that exist between the factors or elements involved, together with the reliability of the scores assigned by users of the device.

Functions to be served by measurement.—With this general inventory of accomplishments before us, then, let us attempt to establish a definite point of departure for a further study of this problem. In doing this our first step is to set up our objectives by making clear just what we wish to serve by a measurement of teaching efficiency. These may be stated as follows:

First, there ought to be a sorting process at the entering door of all teacher-training institutions. Some people, by quality of nature, are foredoomed to failure in teaching, regardless of training and good intentions, and it is a source of waste and disappointment when such people are trained for teaching. A test designed to measure the amounts of the particular traits that are responsible for success might decide, in a few minutes, questions which we have been answering only after several years of experience. To find those traits, define them, and measure them is the task.

Second, we need a means for measuring the progress which is being made in training. Such a device would be valuable, not only as a means of saying when a student has acquired the requisite amount of knowledge and skill to enter the profession, but also, as a basis for directing that training.

Third, we need a test by means of which we may be able to predict the degree of success that is likely to be attained by an

applicant for a teaching position. A test that will, in large measure, replace the long and numerous conferences with superintendents, board committees, etc.

Fourth, we need to measure the efficiency of our teachers in service, in order that: (a) we may supervise more intelligently, (b) promote and remunerate in terms of merit, and (c) replace favoritism and misunderstanding with a rational basis for cooperation.

Fifth, we need to have teachers measure themselves, for the knowledge and professional stimulus they will derive from a thorough analysis of their own work.

Sixth, we need to measure teaching efficiency for the light such measurement will throw upon the teaching and supervisory processes. The writer regards this function of teacher-efficiency tests as a distinct and important function, and not merely as an incident in, or by-product of, the functions stated above. Pedagogical tests have already shown how much clarity can be added to a teaching situation by utilizing the tests for diagnostic as well as for measurement purposes; and there is every reason to believe that teaching tests may help materially toward rationalizing the teacher-training, the teaching, and the supervisory processes.

Criteria for formulating the needed devices.—It is hardly probable that we shall be able to devise any one test or plan of measure that will serve all these purposes. To meet the first, we need, mainly, a test of native endowment, since any glaring lack in the necessary acquired traits would be obvious, either in the general behavior of the applicant or in records of previous training, while any slight defects could still be overcome during training. In addition to this our other aims call for a measure of the teaching process.

Any test designed for use in accomplishing these ends must meet the following requirements: (a) the measurements must be as nearly objective as possible; (b) they must be analytical; (c) there must be no overlapping between factors; (d) the factors must be properly weighted; (e) the plan must not be cumbersome to use; and finally, (f) its validity must be established.

The problem of analyzing general teaching success.—To achieve the above ends within these limitations, we have next to proceed

with our study of the teaching process and of the qualities of teachers themselves. We can hardly hope to be able to speak of the various separate traits or qualities of human nature that are essential to teaching success, nor of the separate functions which a teacher performs, with the same definiteness as that to which we are accustomed in speaking of the organs of the body, or of parts of a machine, or of quantities of size or weight. Yet, if the studies above reviewed argue anything, they argue clearly that we can speak of certain of these factors in teaching success with sufficient definiteness to be clearly understood by others.

Our task is, therefore, that of applying statistical methods to the analysis of "general merit." We must not confuse general with special abilities, and terms used to describe or designate certain functions must be mutually exclusive. For instance, we do not know, except roughly, as shown by Boyce, Ruediger and Strayer, and others, what part "general intelligence" plays.[3] It doubtless affects any value we might assign to "personality," "professional interest," "ability to discipline," etc. One of our first steps, then, if we are to go beyond the present achievement in teacher rating, is to find the significance of general intelligence for "general efficiency" and for all the special factors in "general efficiency." With mental tests available, this step is now only a matter of work. There are doubtless other factors that do not operate independently, and the value of which depends upon their combination with still other factors. To unravel this tangle of special and general factors is our task.

Possible theoretical relationships between factors in "general merit."—We may consider the factors which enter into general teaching success according as they are independent or dependent, equal or unequal,[4] constant or not constant. From this consideration there arise eight possible alternative conditions. Teaching success may consist of factors which are:

1. Independent, equal, and constant.

[3] A considerable amount of carefully collected data in the offices of the department of research of the Oakland school system (Virgil E. Dickson, director) indicates that there is a rather high correlation between "general intelligence" and "teaching success." It is hoped that more exact methods of measuring "teaching success" will soon be available in order that such data as these may make their contribution to this subject.

[4] I. e., equally or unequally potent in contributing to success.

2. Independent and constant, but not equal.
3. Independent and equal, but not constant.
4. Constant and equal, but not independent.
5. Not equal, not independent, and not constant.
6. Independent, but not equal and not constant.
7. Equal, but not independent and not constant.
8. Constant, but not equal and not independent.

If condition No. 1 obtains, we have merely to find out what are the factors, and how much, or how many units of each is present, and then add them together to get "total teaching success." For instance if we have five possible amounts of "ability to discipline," and a given teacher is said to possess three such amounts, then her "general merit" would be raised three points or units by virtue of her possessing this amount of ability to discipline.

If condition No. 2 obtains, it means that the different items contribute varying amounts to "total success." One unit of "ability to discipline" might, for instance, contribute twice as much as one unit of "initiative and self-reliance," in which case the number of units of "ability to discipline" possessed by a teacher would have to be weighted by doubling it. After the items had all been weighted, then their sum would indicate "total teaching success."

If condition No. 3 obtains, then that means that amounts of a given item would not contribute to "general success" in proportion to the number of units of such items possessed. One unit of "skill in discipline" added to "no skill in discipline" would likely not mean anything like as much as one unit added to three units. In other words, up to a certain limit each unit added would add more to "general success" than any previous unit added. It may be true, indeed it almost certainly is true, that not one but many of the factors in teaching success vary in this or the opposite way, or in both ways, in which case a special system of weighting would be necessary for each item before adding to get "total general success."

If condition No. 4 obtains, then we should have to discount certain correlations that might seem to exist between "general success" and each of several separate factors, because a certain group of the separate factors are not independent. They contain

some one or more elements in common. It is almost certain that some of the correlations reported in the studies above reviewed are to be explained in this way. General intelligence may be equally significant for each of all the separate factors involved in "general teaching success." If so, we could deal with the separate factors as if they were independent. If it does not influence them all equally, then here again we have the problem of working out a correct system of weighting, to overcome this lack of independence, e.g., this intercorrelation of separate factors. The value of one factor might vary directly, or inversely as a certain other factor varies. A unit of "ability to discipline" might vary in value as the square, or as the square root of the number of units of "general intelligence."

If any of the other possible combinations of the *characteristics* of the separate factors in "general teaching success" e.g., combinations of *equality*, *constancy*, and *independence*, obtain, then some such system of weighting as has been suggested for combinations 1, 2, 3, and 4 above would have to be worked out.

The next step.—It is obvious that this is not a simple task. Yet, with a fairly clear conception of the theoretical possibilities before us, and with the many correlations already reported for certain factors, by Boyce and others, it would seem desirable as a next step to proceed with our study of the factors already defined by others, trying them out in new combinations, and under different titles, until we shall have found a list of names or criteria by means of which we can recognize the separate factors in "teaching success." Such a study will require time, and the earnest cooperation of a large number of practical school people, but that the end can be attained, that a simple and practical device for measuring teaching efficiency can be worked out, I think there can be no doubt.

Let us follow up our studies of the correlation between "general teaching success" and "training," academic and professional; between "general success" and "general intelligence"; between "general success" and "health"; between "general success" and each of the many separate abilities involved in teaching which have already been worked out, as well as any new abilities or new combination of abilities that can be recognized. We must find out, for instance, whether "ability to discipline" when judged

as one of a total of five factors in "general success" correlates with "general success" in the same way as it does when it is judged as one of a total of twenty separate factors in "general success." And, as suggested above, we must find out the correlation of "general intelligence" with each of the special abilities or factors in "general success."

It is not the purpose here to imply that the several devices now available for measuring teaching efficiency are wholly unsatisfactory. Far from it. The plans of Elliott, Boyce, Landsittel, Rugg, and others have been used with fair success. But such systems as these are not in general use. The idea of this sort of measurement is not very widely accepted as yet, and the practice is still further short of being common over the country. To bring together the fruits of our study to date, and to use their contribution as a starting point for an extended study in which many practical school people will cooperate is our present need and should be our next step.

BIBLIOGRAPHY

1. Anderson, W. N. "The selection of teachers," *Educational Administration and Supervision*, 3:83–90, February, 1917.
2. Bird, Grace E. "Teachers' estimates of supervisors," *School and Society*, 5:717–20, June 16, 1917.
3. Book, W. F. "The high school teacher from the pupils' point of view," *Pedagogical Seminary*, 12:239–88, September, 1905.
4. Boyce, A. C. "A method for guiding and controlling the judging of teaching efficiency," *School Review Monographs*, No. 6, pp. 71–82, 1915.
5. Boyce, A. C. "Methods for measuring teachers' efficiency," *Fourteenth Yearbook of the National Society for the Study of Education*, Part II, 1915. (See also *Journal of Educational Psychology*, 3:144, March, 1912.)
6. Bradley, J. H. "A study of the relative importance of the qualities of a teacher and her teaching in their relation to general merit," *Educational Administration and Supervision*, 4:358–63, September, 1918.
7. Buellesfield, Henry. "Causes of failures among teachers," *Educational Administration and Supervision*, 1:439–45, September, 1915.
8. Clapp, F. L. "Scholarship in relation to teaching efficiency," *School Review Monographs*, No. 6, pp. 64–70, 1915.
9. Clark, R. C. "A scale for measuring teachers," *American School Board Journal*, 62:39–40, February, 1921.
10. Coffman, L. D. "Committee on rating, placing and promotion of teachers," *School Review Monographs*, No. 6, pp. 61–63, 1915.
11. Coffman, L. D. "The rating of teachers in service," *School Review Monographs*, No. 5, pp. 13–24, 1914.

12. Colvin, Stephen S. "The most common faults of beginning high school teachers," *School and Society*, 7:451–59, April 20, 1918. (Also in the *Eighteenth Yearbook of the National Society for the Study of Education*, Part I, pp. 262–72, 1919.)
13. Connor, W. L. "A new method of rating teachers," *Journal of Educational Research*, 1:338–58, May, 1920.
14. Cook, William A. "Uniform standards for judging teachers in South Dakota," *Educational Administration and Supervision*, 8:1–11, January, 1921.
15. Courtis, S. A. *Manual of directions for making efficiency card ratings*. Detroit, Michigan: Board of Education, May, 1920.
16. Cranor, Katherine T. "A self-scoring card for supervisors as an aid to efficiency in school work," *Educational Administration and Supervision*, 7:91–102, February, 1921.
17. Elliott, E. C. *Provisional plan for the measure of merit of teachers*. Madison; Wisconsin State Department of Education, 1912.
18. Elliott, E. C. "How shall the merit of teachers be tested and recorded," *Educational Administration and Supervision*, 1:291–99, May, 1915.
19. Fichandler, A. "A study in self-appraisal," *School and Society*, 4:1000–2, December 30, 1916. (See criticisms of this article by Taylor and Myers, pages 173 and 322 respectively of *School and Society*, v. 5, February 10-March 17, 1917.
20. Fordyce, Chas. "Notes on the correlations between general teaching power and some specific teaching qualities," *Eighteenth Yearbook of the national Society for the Study of Education*, Part I, pp. 349–51, 1919.
21. Foster, Frederick M. "A score card for rural teachers," *School and Society*, 12:131–32, August 21, 1920.
22. Gray, William S. "Rating scales, self-analysis and the improvement of teaching," *School Review*, 29:49–57, January, 1921.
23. Hickman, James. "A measuring scale for teachers in service," *American School Board Journal*, 52:43–44, April, 1916.
24. Hill, L. B. "A card for rating practice teachers and teachers in service," *Report of the Society of College Teachers of Education*, 1920.
25. Johnston, Joseph H. "Scientific supervision of teaching," *School and Society*, 5:181–88, February 17, 1917.
26. Johnston, Joseph H. "Teacher rating in large cities," *School Review*, 24:641–47, November, 1916.
27. Jones, Edward S. "A suggestion for teacher measurement," *School and Society*, 6:321–22, September 15, 1917.
28. Kent, R. A. "What should teacher rating schemes seek to measure," *Journal of Educational Research*, 2.802–7, December 1920.
29. Kratz, H. E *Studies and observations in the schoolroom*, Chicago: Educational Publishing Company, 1907, chapter 5.
30. Landsittel F. C. "Evaluation of merit in high schoo teachers," *School and Society*, 6:774–80. December, 1917.
31. Landsittel, F. C. "A score card method of teacher rating" *Educationa Administration and Supervision*, 4:297–309, June, 1918.
32. Littler, Sherman. "Causes of failure among elementary school teachers," *School and Home Education*. 33:255–56, March, 1914.
33. Moody, Floyd E. "Correlation of the professional training with the teaching success of normal school graduates," *School Review*, 26:180-81, March, 1918.

34. MORTON, ROBERT L. "Qualities of merit in secondary teachers," *Educational Administration and Supervision*, 5:225–38, May-June, 1919.

35. Moses, Cleda. "Why high school teachers fail," *School and Home Education*, 33:166–69, January, 1914.

36. PATTERSON, HERBERT. "How can the ability of student teachers be measured," *Educational Administration and Supervision*, 6:215–19, April, 1920.

37. PITTENGER, B. F. "Problems of teacher measurement," *Journal of Educational Psychology*, 7:103–10, February, 1917.

38. PUSEY, EDWARD D. "Testing grade teachers for efficiency," *Journal of Proceedings and Addresses of the National Education Association*, 1914, pp. 283–86.

39. RISLEY, JAMES H. "Selecting teachers and grading their efficiency," *American School Board Journal*, 49:11, 69, September, 1914.

40. RUEDIGER, WILLIAM C. *Agencies for the improvement of teachers in service.* (United States Bureau of Education Bulletin, 1911, No. 3.)

41. RUEDIGER, WILLIAM C. and STRAYER G. D. "The quality of merit in teachers," *Journal of Educational Psychology*, 1:272–78, May, 1910.

42. RUGG, H. O. "Self-improvement of teachers through self-rating: a new scale for rating teachers' efficiency," *Elementary School Journal*, 20:670–84, May, 1920.

43. SANTEE, A. M. "The organization and administration of practice-teaching in state normal schools," *School and Home Education*, 37:8–13, September, 1917.

44. SHIDELER, S. E. "Qualifications, salary and tenure of the teachers in the commissioned high schools of Indiana," *School Review*, 21:446–60, September, 1913.

45. SHOWALTER, B. R. "A score card for rural teachers," *School and Society*, 12:200, September 11, 1920.

46. SPAULDING, F. E. *What Minneapolis teachers think of the superintendent, supervisors and principals.* Minneapolis: Board of Education, 1917.

47. SPRAGUE, H. A. "Score-card for rating student-teachers in training and practice," *Pedagogical Seminary*, 24:72–80, March, 1917.

48. "A teachers' rating card," *Elementary School Journal*, 20:723–24, June, 1920.

49. THOMPSON, FRANK E. "The rating, placing, and promotion of teachers," *School Review Monographs*, No. 5, pp. 1–11, 1914.

50. *Teacher Personality* (score card). Institute for Public Service, New York City.

51. TWISS, GEORGE R. "A plan for rating the teachers in a school system," *School and Society*, 9:748–56, June 21, 1919.

52. "Who shall receive increases in salary?" *Elementary School Journal*, 20:492–94, March, 1920.

53. WAGNER, C. A. "The construction of a teacher rating scale," *Elementary School Journal*, 21:361–66, January, 1921.

54. WITHAM, E. C. "School measurement," *Journal of Educational Psychology*, 5:571–88, December, 1914.

55. WITHAM, E. C. "School and teacher measurement," *Journal of Educational Psychology*, 5:267–78, May, 1914.

ON THE NEW PLAN OF ADMITTING STUDENTS AT COLUMBIA UNIVERSITY[1]

Transmitted by
EDWARD L. THORNDIKE
Teachers College, Columbia University

THE FIRST QUOTATION: DEAN HAWKES

Fortunately it is possible to determine with scientific accuracy whether or not the mental test is a useful addition to our academic machinery. If it turns out, during a series of years, that the correlation between the marks received on the mental tests and the collegiate work of the students is distinctly higher than the correlation between the results of other types of entrance examinations and the college work, it would seem to be clear that the new plan of admission affords the best index that we have of the ability of a boy to carry college work. The correlation between the work of the entire freshman year for the students who entered by the new plan and their marks on the mental test is +0.65. The most reliable data available indicate that the highest correlation that can be expected between the work of the freshman year and the results of the usual college entrance examinations is about +0.45. This latter figure has been obtained not only from a statistical study of our own freshmen but from similar studies in another institution. Although it is too early to make a final statement regarding the matter, every indication points to the mental test as a most useful addition to our machinery of admission. It must be kept in mind that the group of students who are admitted to college under the new plan are very carefully winnowed before they are authorized to take the mental test. The correlations obtained should, therefore, be interpreted as referring to the new plan of admission as a whole rather than to the mental test alone.

[1] At the public meeting of the National Association of Directors of Educational Research at Atlantic City, March 30, 1921, Professor Thorndike discussed the new plan of admitting students to Columbia University. He based his discussion in part on these two quotations. The first is by Doctor Herbert E. Hawkes, Dean of Columbia College, and the second is by Doctor Adam L. Jones, Director of University Admissions.

In addition to the use of the results of the mental test in admission to college, they have been most helpful in my office as an aid in arriving at a diagnosis of academic maladies. A boy who has a poor academic record and a low mental-test grade generally needs very different treatment from the student whose record is poor but whose mental-test mark is high. And in several cases the mental test has afforded the clue which has enabled my office in cooperation with the university physician so to advise the boy that he has not only escaped being dropped, but has become an excellent academic citizen.

The wise use of a new instrument like the mental test requires constant caution and scrupulous checking, but its apparent possibilities for usefulness are so fundamental and far reaching that a careful and scientific study of its significance is one of the important tasks of the next few years.

The Second Quotation: Director Jones

It will be remembered that the new method permits students whose school and character records are satisfactory to us to substitute this examination for the entrance examinations. Candidates still have the option of entering by the old method, and they still have the privilege of substituting for the entrance examinations certain of the examinations given in schools by the New York State Department of Education. The records of those electing the older method of admission were scrutinized with the greatest care. The requirements were very strictly enforced. Those entering by the New York State examinations were required to have 70 percent or higher in each subject and admission with conditions was allowed only in the case of those whose outstanding excellence more than made good a technical deficiency.

This requirement was most strictly enforced in the case of those who came from the best high schools where the instruction was of the highest quality, and where in consequence there was least excuse for a doubtful record. A student with a poor record from a good school is usually a bad risk. Students coming from small or poorly equipped high schools were treated with greater leniency. A good student may fail to make a first-rate record

in a poor school. Where there was room for doubt, the candidate was required to take the mental test.

By far the most significant group so far as our system of admission is concerned was the group entering by the new method.

It was expected that this method would appeal strongly to enterprising and alert young men in places in which the New York State examinations are not given and where college entrance examinations are less well or favorably known than in schools from which our student body has usually been chiefly drawn. The expectation was more than realized, the number of candidates for the freshman class from a distance was much greater than in the past, and those who applied under these conditions were generally successful in the test. It should be remembered that only those whose school records and character records were entirely satisfactory were allowed to enter by the new method. They constituted, therefore, a picked group.

That they were a picked group is evident not only from the records which they presented for admission, but also from the records which they made in college. They have done remarkably well. There were, of course, borderline cases and a number of these have not turned out well, but in the group as a whole the failures were very few. Most of those who failed were students whose scores in the psychological examination were relatively low, but whose cases seemed to possess sufficient merit to warrant their being given a trial. Even in this group, most justified their admission to college. There were a very few with relatively high scores whose records in college were not wholly satisfactory. Careful examination showed that their failure to make first-rate records was due not to lack of intelligence nor to faulty preparation, but to failure to divide their time and energies properly among the many demands which come to the college student.

Meetings of instructors of freshmen regularly follow the making up of the mid-term and term records for the purpose of considering the cases of students whose records are unsatisfactory. In the meeting last November only two of the students among more than sixty whose records were unsatisfactory had made high scores in the psychological examination. The testimony of their instructors was unanimously to the effect that both students were fully able to do good college work. It appeared, however,

that one of them had been devoting too great an amount of attention to athletics and other extra-curricular activities, while the other had been taking undue advantage of his first opportunity to become acquainted with a great city.

Later study has shown that with remarkably few exceptions the higher a student's score in the psychological examination, the better his record in college.

Several studies have been made in the course of the year bearing upon this point. Some of these have been made by Mr. Harold K. Chadwick of this office. Others have been prepared by Mr. Ben D. Wood of the Department of Psychology. In one of Mr. Chadwick's studies he considered one hundred and eighty men made up of groups of ten, each group having psychological examination grades lying within two degrees in the scale from 70 to 106. The grades of the highest ten covered a range of three degrees since there were fewer than ten within two degrees. With this exception each two degrees in the scale from 70 to 106 was represented by a group of ten, the first ten alphabetically being taken in each case. The total amount of work in points done by each group was plotted by grades. The result for three typical low, middle, and high groups is given below:

Group Scores Between	Amount of Work in Points at Each Given Grade											
	F	D−	D	D+	C−	C	C+	B−	B	B+	A−	A
70 and 72	33	0	13	2	12	26	8	6	21	4	0	0
86 and 88	9	3	14	4	10	18	25	8	36	7	3	3
104 and 106	0	0	9	0	7	24	18	19	54	16	3	16

The progressive decrease of low grades and increase of high grades as we go from the low to the high groups is significant. It was found to hold good very generally throughout the eighteen groups though there were a few exceptions.

No group of men with psychological examination grades below 78 received A's, only one group below 84 received A's. No group above 100 received F's, only one above 95 did so.

In another study the work of groups covering five degrees was compared and the dividing mark, between the upper and upper middle quartiles, the lower middle and upper middle quartiles and the lower middle and lower quartiles were studied with the following results:

Group	Dividing Line for Lowest Quartile	Dividing Line Between Two Middle Quartiles	Dividing Line for Highest Quartiles
70– 75	F	C–	C+
75– 80	D	C–	C+
80– 85	D+	C	B
85– 90	C–	C	B–
90– 95	C	B–	B
95–100	C	B–	B
100–105	C	C+	B
105–	C+	B	B+

The men with higher grades did a larger amount of high-grade work and a smaller amount of low-grade work though the groups from 90 to 105 were practically the same so far as this study shows. It will be seen for example, that for the men with a psychological examination grade of 105 "C" marked the division between the lowest fourth and the remaining work while for those with a grade below 80 "C" marked the division between the highest quarter and the remaining work; three-fourths of the work of the first was as high as the highest fourth of the work of the second group.

Another study of the work of students in the second session gave a number of striking results, all of which went to prove the progressive superiority of the higher groups in the order of their grades. The following is typical:

GROUP OF FIFTY FOR EACH TEN DEGREES OF THE SCALE

Grades	Number Doing "B" Work or Better		Percent of All "B" Men[1]
	Individuals Out of Group of Fifty	Percent of Group	
60– 70	2	4	3¾
70– 80	4	8	7½
80– 90	7	14	13¼
90–100	15	30	28¼
100–120	25	50	47¼

[1] "B men" were 21.2 percent of the 250 men.

It should not be supposed, however, that the psychological examination alone would give ideal results particularly when the scores fall below 80. The range from 60 to 70 was regarded as extremely doubtful. While the complete school records of all candidates were carefully examined, especially close attention was given to those whose scores were between 60 and 70, and only the most promising were admitted. The records show that this group did work practically equal to that for which the range was 70 to 80 and whose records had not been quite so carefully weighed.

The psychological examination alone would not be a fully satisfactory means of selecting students, but there has been no thought of using it without the student's complete previous record. Ordinarily the candidate's school record must show the completion of a school course covering the requisite entrance subjects with grades 10 percent or more above the school's passing mark. His personal record must show acceptable mental and moral qualities. Occasionally a student of especial promise with a record which is doubtful in certain particulars may be allowed to take the examination with the requirement that he pass with a very high grade.

It will be recalled that students who elect to enter by the old method take the psychological examination for purposes of record. This makes it possible each year to test the results of the psychological examination for practically every student in college.

A preliminary comparison of the relation between college record on the one hand and school record, entrance examination, regents' examinations, and psychological examination on the other was made at the close of the winter session by Mr. Ben D. Wood of the Department of Psychology. The results are significant. Among the students admitted by the college entrance examinations a good many doubtful cases were included. The correlation between their examinations and their college records was +0.43 which is reasonably satisfactory. The correlation between school record and college record was +0.45. Those entering by regents' examinations were very carefully selected. The correlation between their examination and their college records was +0.57, while the correlation between psychological examination and college record was +0.59, a highly satisfactory result. This was for the first half-year only.

A similar study for the work of the whole year shows a correlation between mental test and college record of $+0.60$ which was remarkably good. The correlation for the other examinations and for the school record has not yet been worked out. It should be remembered that there are many factors other than intelligence which determine a student's standing and that the psychological examination is not supposed to measure them.

The operation of the new system will be watched with the greatest care and no opportunity for checking its results or improving the methods of using it will be lost. It may still be an experiment, but it is certainly not a doubtful one.

A METHOD OF EQUALIZING THE RATING OF TEACHERS

LEE BYRNE
Supervisor of High-School Instruction, Dallas, Texas

It is well known that ordinary marks given to pupils by different teachers show considerable variation due to the teacher's personal equation. Some teachers tend to mark very high, others very low. Hence a number of plans have been worked out to secure greater consistency in pupils' marks within a particular school.

A still greater degree of variation may be found in the rating of teachers by principals in different buildings in the same school system. In marking pupils instructors are likely to exchange opinions and information regarding individuals, and this influence produces a tendency toward consistency. The rating of teachers, however, is likely to be made by each principal quite independently. Moreover, no teacher has more than one principal, whereas each pupil in a departmentalized school may have several teachers.

Inconsistency on the part of principals in the rating of groups of teachers may be described statistically under two points: (1) differences in the general *level* of the ratings made, and (2) differences in the *spread* of the ratings. The first has to do with the central tendency or average, and the second with the dispersion.

To eliminate the differences in the general *level* of ratings made by different principals, it is only necessary (assuming the ratings to be expressed numerically) to calculate the average for each school and for all schools combined and then to move each school average up or down to make it coincide with the average for the whole system.

This procedure assumes that the real average of teaching ability in each school is the same and equal to the average for the whole system. This may not be true, though it is likely to be approximately so. The fact that the level of teaching ability in a given school appears unusually high or low is more likely to be due to liberal or rigorous rating than to an actual condition. Moreover, if the level of teaching ability is really markedly above

or below the general level, this condition is a defect which ought in justice to be remedied.

To eliminate the differences in the *spread* of the ratings, the standard deviation may be found for each school and for the system as a whole; then each teacher's position in the original rating report, in terms of the standard deviation of that school, may be reexpressed in terms of the standard deviation of the whole system from the average of the whole system.

The usual method of attack on the problem of pupils' marks is to devise a scheme of control by which teachers will distribute their marks in certain prearranged proportions. The method of equalization of ratings just indicated consists in taking the raw data as submitted by the principals and using statistical calculations in the central office to reduce all the reports to a uniform level and variability. It would be unduly laborious to apply this method to pupils' marks; but it is quite feasible and advantageous as a method of organizing crude ratings of teachers and of giving them final form. It can be applied to the current ratings of a particular year, and it will also be found very useful as a method of organizing and making comparable the ratings of a series of years past. Moreover, it can be used for this purpose even when different schemes of rating have been employed—provided only that the ratings are numerical or convertible into numbers.

An illustration is given here of the method of reducing crude ratings to uniformity of level and dispersion. In the particular rating scheme reported each teacher is marked on ten points, each with five degrees of quality, these degrees being represented with numerical values of 6, 8, 10, 12, and 14. Thus a teacher's minimum possible numerical rating would be 60 and the maximum would be 140. The method of equalization is equally applicable to any other rating scheme so long as numerical values are used.

The first step is to tabulate the crude ratings and calculate for each school and for the group of schools the arithmetic mean (or average) and the standard deviation. This has been done for a system consisting of seven schools with results as shown in Table I.

TABLE I. CENTRAL TENDENCIES AND DISPERSIONS OF RATINGS

	Schools							Entire System
	1	2	3	4	5	6	7	
Number of teachers.....	16	17	14	23	10	12	33	125
Arithmetic mean........	107.1	121.6	108.4	114.8	110.6	120.2	123.4	116.5
Standard deviation......	8.0	13.5	7.9	10.7	6.9	11.5	9.0	11.7

The second step is to take the arithmetic means and standard deviations found and construct a conversion table which will enable us to find for any crude rating the revised or standardized rating desired. The conversion table is illustrated in Table II.

To construct a conversion table first place the values for the arithmetic means in the row marked 0.0σ. Then from this point measure upward and downward various intervals from $+2.5\sigma$ to -2.5σ. A complete set of values for any number of subdivisions may be filled in, but in practice it is sufficient to put down a framework of values—e.g., at intervals of 0.5σ as in Table II—from which all individual marks can readily be calculated. It is easy to see for any crude rating what the corresponding standardized rating would be in any proposed scale. Theoretically the crude values would be transformed into those of the "Composite" scale, based on the array of ratings from all the schools. In practice it will be less troublesome and equally serviceable to transform to a conveniently numbered scale such as "B" or "D," which approximates the conditions of the composite. Or if preferred a scale stopping at 100 can be employed, such as "A."[1] In the last column of Table II five general subdivisions of quality are indicated; but any other number can readily be employed.

According to Table II, a teacher in School 1 who was rated 123 would be entitled to a rating of 140 (to the nearest whole number) on the composite scale, 143 on scale D, 130 on scale B, and 95 on scale A. On the other hand, a teacher in School 2, due to the more liberal rating prevailing in that school, who was

[1] However it is advisable to avoid the use at any time of a marking scale which stops at 100 percent; the preconceived idea that marks must fall between 80 and 100 or even 90 and 100 is hard to resist.

TABLE II. FOR CONVERTING CRUDE RATINGS INTO A UNIFORM SCALE

	Schools							Scale "C" (Composite)	Scale "D" (75–150)	Scale "B" (85–135)	Scale "A" (50–100)	General Subdivision of Quality
	1	2	3	4	5	6	7					
$\sigma =$	8.0	13.5	7.9	10.7	6.9	11.5	9.0	11.7	15	10	10	
$0.5\sigma =$	4.0	6.75	3.95	5.33	3.43	5.74	4.49	5.87	7.5	5	5	
$0.1\sigma =$	0.8	1.35	0.79	1.07	0.69	1.15	0.9	1.17	1.5	1	1	
2.5σ	127.1	155.4	128.2	141.4	127.8	148.9	145.9	145.8	150	135	100	A
2.0σ	123.1	148.6	124.2	136.2	124.4	143.2	141.4	139.9	142.5	130	95	"Very Much
1.5σ	119.1	141.9	120.3	130.8	120.9	137.4	136.9	134.1	135	125	90	Above Average"
1.0σ	115.1	135.1	116.3	125.5	117.5	131.7	132.4	128.2	127.5	120	85	B
0.5σ	111.1	128.4	112.4	120.1	114.0	125.9	127.9	122.4	120	115	80	"Above Average"
0.0σ	107.1	121.6	108.4	114.8	110.6	120.2	123.4	116.5	112.5	110	75	C
-0.5σ	103.1	114.9	104.5	109.4	107.1	114.4	118.9	110.7	105	105	70	"Average"
-1.0σ	99.1	108.1	100.5	104.1	103.7	108.7	114.4	104.8	97.5	100	65	D
-1.5σ	95.1	101.4	96.6	98.7	100.2	102.9	109.9	99.0	90	95	60	"Below Average"
-2.0σ	91.1	94.6	92.6	93.4	96.8	97.2	105.4	93.1	82.5	90	55	E
-2.5σ	87.1	87.8	88.7	88.1	93.4	91.5	100.9	87.2	75	85	50	"Very Much Below Average"

rated 149 would be entitled to no higher ratings on the uniform scales. Any rating obtained in any school can be read off on one of the uniform scales. Interpolation is, of course, necessary especially when, as in Table II, intervals are as large as 0.5σ.

The result of using such a conversion table in order to standardize the ratings may be seen in Table III. Under the caption "Crude Scale" the teachers are entered according to the rating made by the principals of the several schools. It is from these ratings that the arithmetic means and standard deviations shown in Table I were computed. Under the caption "Scale B" the same ratings are shown when converted into the units of Scale B. The larger differences both in general level and in distribution—differences probably due mainly to different ways of rating—are greatly reduced when the uniform scale is used.

TABLE III. DISTRIBUTION OF CRUDE RATINGS AND STANDARDIZED RATINGS ACCORDING TO SCALE "B"

Rating	Crude Scale								Scale B							
	Schools							All Schools	Schools							All Schools
	1	2	3	4	5	6	7		1	2	3	4	5	6	7	
138	...	1	...	...	...	...	1	2	...	...	...	...	...	...	.....	....
136	...	1	...	1	...	1	1	4	...	...	...	...	...	...	.....	....
134	...	1	...	...	...	...	...	1	...	...	...	...	...	...	.....	....
132	...	1	...	...	...	1	7	9	...	...	...	...	...	...	.....	....
130	...	1	...	1	...	2	2	6	...	...	...	1	1	...	...	2
128	...	1	...	2	...	1	2	6	...	...	...	...	...	...	.....	....
126	...	3	...	1	...	...	5	9	3	...	...	...	...	...	1	4
124	...	...	...	2	1	1	1	5	...	...	...	1	...	1	1	3
122	...	1	...	1	...	...	2	4	...	1	2	2	...	...	...	5
120	3	1	...	...	...	...	3	7	1 ...	1 ...	2	1 ...	...	1	7	13
118	...	...	2	1	...	1	2	6	...	1 1	...	2 ...	2	2 ...	...	8

Rating	Crude Scale								Scale B							
	Schools							All Schools	Schools							All Schools
	1	2	3	4	5	6	7		1	2	3	4	5	6	7	
116	1	3	2	1	2	2	...	11	...	1	3 ...	1 ...	...	1 ...	2 2	10
114	...	1	3	1	1	...	1	7	...	1 ...	...	...	1 ...	...	...	2
112	...	...	...	3	1	...	1	5	2 ...	3 ...	...	1 ...	1	1 ...	5 ...	13
110	2	1	...	2	1	...	3	9	1 ...	1	1	1 ...	...	...	1 ...	5
108	1	...	1	1	...	2	...	5	1 ...	1 ...	...	1 3	1 ...	1	2	10
106	1	...	1	2	1	...	...	5	2	3	1 ...	2	...	2	3	13
104	2	...	...	...	2	...	2	6	3	1	...	1	...	...	2	7
102	3	...	...	1	...	...	...	4	...	1	...	2	1 ...	...	...	4
100	...	...	3	2	1	...	...	6	...	...	3	...	2 ...	...	1	6
98	1	...	1	...	...	...	...	2	1 ...	...	...	1	...	2 ...	...	4
96	2	...	1	1	...	1	...	5	2	...	1 ...	2	...	...	1 3	9
94	...	...	...	...	...	...	...	......	...	...	1	...	1 ...	...	...	2
92	...	...	...	...	...	...	...	......	...	...	...	1 ...	...	...	...	1
90																0
88	...	...	...	...	...	...	...	......	...	...	...	...	...	1 ...	2	3
78	...	1	...	...	...	...	...	1	...	..1	...	...	...	...	...	1
Total..	16	17	14	23	10	12	33	125	16	17	14	23	10	12	33	125

Figure 1 possibly shows the character of the situation more clearly. In the crude ratings Schools 2, 4, and 7 have a large proportion of their marks higher than the maximum attained in Schools 1 and 3. But in Scale B it is evident that there is a fairly balanced distribution.

The writer would not claim that the assumptions on which this suggested method of organizing crude ratings is based correspond exactly to the truth, but he would submit that if principals are to file teachers' ratings year after year, these records can be rendered far more reliable, significant, and valuable if they are subjected to some statistical treatment such as the one illustrated.

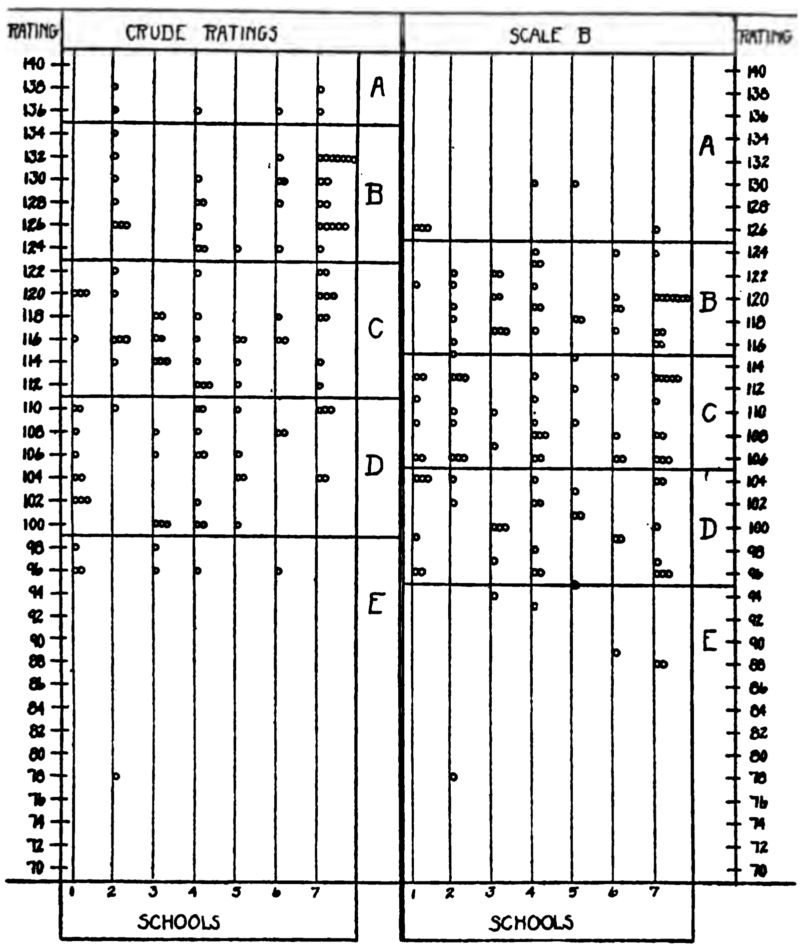


FIGURE 1. COMPARISON OF CRUDE RATINGS AND RATINGS BY UNIFORM SCALE B. DATA FROM TABLE III

COOPERATIVE CHEMISTRY TESTS[1]

SETH HAYES
East Technical High School, Cleveland, Ohio

BASIS OF THE WORK

The chemistry teachers of Cleveland have been cooperating for about a year in an effort to formulate questions which would be a measure of the information in chemistry possessed by their pupils. Questions upon limited portions of the semester's assignment were sent out from time to time to the different teachers for use in their classes. A record was made for each question on the basis of returns from the teachers. These questions were gleaned from different sources, many being submitted by the teachers themselves. As a consequence, there are on hand some five hundred questions based on the textbook in use (McPherson and Henderson), which have been tried out in three or more schools and have been attempted by at least one hundred pupils. It is intended that this work shall be followed up until about eight hundred questions have been standardized for the subject as presented in Cleveland.

The cooperative feature of these tests is in itself of the greatest value. We are trying to prepare a measure for use in our own work. Already good effects have come from the scrutiny which we have individually given to our work as a result of this united effort. New projects which have developed directly or indirectly from this work await their turn for consideration and treatment.

PREPARATION OF QUESTIONS

The questions prepared for these tests called for equations or single-word answers. It was the aim that they should conform to the requirements for such tests as laid down by Dr. J. Crosby Chapman.[2]

[1] In connection with this study, appreciation is due and very gladly expressed to the teachers for their aid and interest, and for their kindly adaptation to the conditions of the tests, which often interfered with their customary individual procedure; also to the headquarters staff, especially Mr. Welles, for sympathetic support. Particular thanks are due to Dr. Chapman, at whose suggestion the work was undertaken and by whose advice it was carried on.

[2] Chapman, J. Crosby. "The measurement of physics information," *School Review*, 27:748–56, December, 1919.

The committee which prepared these tests based their selection on about two hundred special questions which had been submitted by the teachers for this purpose. In order to place all of the pupils on essentially the same basis none of these questions was drawn from those used in previous (preliminary) tests, which had been tried out to a certain extent in the various schools and classes of the city. The members of the committee, the writer excepted, are not teaching the subject at present, and therefore had no pupils to undergo the test. None of the pupils of the writer was given the tests. The questions as given, although based on those submitted, were edited so as to remove the influence of the special presentation of the individual teachers. The genesis of the thirty-nine questions used is as follows: (a) Preliminary experimental tests were given from time to time, consisting of questions from various sources; (b) Tentative valuations were placed on such of these questions as seemed justifiable; (c) Numerous other questions for the purpose of these tests were submitted after we had had the experience with the preliminary work; (d) These were edited and used.

Characteristics of the Tests

These tests are intended to be rapid-fire, and to call forth quick and accurate thinking by the pupils. They are given in the following manner: Each question is read, reread and immediately answered by the pupils; then the papers are marked, either by the teacher alone or by the aid of pupils. These tests offer the advantage of taking but little time to give (35 minutes) and to grade, they pin the pupils down to exact answers, and they require no special material.

This style of test is in no way to be considered as a substitute for the usual examination, but it has many advantages for the measurement of information. The final test questions were constructed to measure not only the information possessed, but reasoning and constructive ability as well. To do this, a question or two was included in each set which was purposely made too difficult for use in the usual prorata test; but which served the purpose of indicating the very best thinkers in chemistry in the classes. The frequent use of this style of test cultivates in the pupils speed and the ability to differentiate values, not only during the tests, but in regular recitations as well.

Directions to Teachers

The following directions were given to the teachers for conducting the tests.

Method of Administering the Test.

1. Before starting, instruct the pupils to fill in the blanks at the head of the "Answer Sheet," Figure 1, except the "Total Right."

FIGURE 1. ANSWER SHEET

Name............................. School...............................
Date.....................Topic.................... Total Right...........

No.	Result	Answers	Answers	Result	No.
1					11
2					12
3					13
4					14
5					15
6					16
7					17
8					18
9					19
10					20

2. Instruct the pupils to write their answers in the large spaces which are headed, "Answers." Each answer should be in the space which corresponds to the number of the question given. The answers should be in *one word*, if possible. Under no circumstances should any writing be done in the "Result" spaces.

3. Read each question to the class slowly and clearly; then repeat it immediately. Watch your pupils to judge the time needed in answering the question in hand. Do not change, explain, or omit any of the questions.

4. Before collecting the papers, reread the questions, allowing corrections to be made.

Method of Marking the Papers

5. When the time is up on the last question, stop all work. Mark the correctness of the answers in the "Result" spaces, Figure 1, by means of a large check if correct, and by a large cross if incorrect.

6. Count the number of correct answers on the sheet and enter the number in the space labeled "Total Right."

Use to be Made of the Final Scores of Each Pupil

7. On the "Summary Sheet," Figure 2, enter the "Scores of the Questions." This will indicate the number of pupils in each class who answered the indicated question correctly.

8. Also obtain and enter the "Scores of Pupils." This will indicate the number of questions answered correctly by each pupil in each class.

9. The balance of the sheet will be filled out by the person who is conducting the experiment.

(NOTE: The "Answer Sheet," Figure 1, was prepared for convenience in marking. These sheets can be easily handled in one of two ways: (a) If stacked they can be fingered over rapidly, the answers to two or three questions being marked at each going-over; or (b) The sheets can be spread out laterally so as to expose either the right or left halves of all the sheets and then one-half of all of the answers can be most rapidly marked. The "Summary Sheet," Figure 2, was prepared for the sake of uniformity in the making of reports.)

THE TESTS: QUESTIONS AND ANSWERS

The questions are here arranged in order of increasing difficulty as shown by the results of the tests. The question numbers indicate the original order.

FIRST SEMESTER TEST

(Ground covered: McPherson & Henderson, chapters I–XVII)

VALUES	QUESTIONS	ANSWERS
1.100	2. How is pure water usually prepared in the laboratory from impure water?	Distillation
1.752	1. If a piece of magnesium is burned, how does the weight of the resulting solid compare with the weight of the original piece of magnesium?	Greater
2.469	6. In preparing oxygen from potassium chlorate, manganese dioxide is added to the chlorate. What kind of an agent is the manganese dioxide in this reaction?	Catalytic, Catalyzer

2.508	11. In putting out a fire, one of the three factors of combustion must be removed. In the use of a hand fire-extinguisher, which factor is removed?	Supporter
2.586	12. What is formed in addition to nitric oxide and oxygen when nitric acid decomposes in its usual manner as an oxidizer?	Water, H_2O
2.624	13. What is hydrated ammonia?	Ammonium hydrate, Ammonium hydroxide, Aqua ammonia, NH_4OH
2.926	9. What compound is always formed by neutralization?	Water, H_2O
2.963	8. Unless adjustments are made by the pilot, a balloon tends to descend towards evening and to rise toward midday. By the aid of what gas law can this be explained?	Charles', Gay-Lussac's
3.000	3. When hydrogen is passed over hot copper oxide a chemical change takes place. What is the action of the copper oxide?	Oxidizing agent, Oxidized the hydrogen
3.037	10. If nitrogen were prepared by burning out the oxygen from some air, what very inactive element would make up about 1/79 of the unconsumed gases?	Argon, A
3.224	5. Write a molecular reaction expressing the oxidation of an element which shows that water is a product of combustion?	$2H_2+O_2=2H_2O$
3.453	15. The valence of the element "Y" is –4 and that of the element "X" is +3. Write the formula of the binary compound which these elements could form.	X_4Y_3
3.531	17. Ammonia will escape from a bottle of ammonium hydroxide as long as the bottle is unstoppered. If tightly stoppered, equilibrium is soon established. Express this equilibrium in a reversible reaction.	$NH_4OH \rightleftarrows NH_3+H_2O$
3.531	7. From what compound can pure nitrogen be prepared by heat?	Ammonium nitrate, NH_4NO_3
3.652	14. Complete and balance the following reaction: $Ca(OH)_2+H_3PO_4=$	$3Ca(OH)_2+2H_3PO_4=Ca_3(PO_4)_2+6H_2O$ *or* $Ca(OH)_2+H_3PO_4=CaHPO_4+2H_2O$

3.693	18. Write the equation for the complete action between copper and hot concentrated sulphuric acid	$Cu + H_2SO_4 = CuSO_4 + 2H_2O + SO_2$
3.735	4. Write the equation for the reaction which takes place when nitric acid is prepared at a moderately low temperature.	$NaNO_3 + H_2SO_4 = NaHSO_4 + HNO_3$
3.820	16. What do we call the action of water upon salts by which a base and an acid are formed?	Hydrolysis
4.900	19. Write the equation for the action which takes place when hydrogen sulphide is passed into concentrated nitric acid, liberating sulphur.	$2HNO_3 + 3H_2S = 4H_2O + 2NO + 3S$

SECOND SEMESTER TEST

(Ground covered: McPherson and Henderson, chapters XVIII–XLII)

Values	Questions	Answers
1.012	2. What process beside the acid-Bessemer process is generally used in the United States in producing the steel of commerce?	Basic Open-hearth
2.046	10. When iron ores are mixed with a suitable flux and are reduced with coke, what is the main product?	Cast (Pig) Iron
2.136	5. Write the equation for the reaction which takes place between chlorine and water.	$2H_2O + Cl_2 = 4HCl + O_2$
2.136	15. What element because of its affinity for oxygen is most generally used in metallurgy as a reducer?	Carbon, C
2.180	6. What metal which is lighter than water will decompose water and set free half of the hydrogen without the hydrogen taking fire if the water is cold?	Sodium, Na
2.307	13. The insoluble soap which gathers on the top of hard water in washing is apt to be the salt of what metal?	Calcium, Ca (Magnesium, Mg.)
2.388	11. We breathe out carbon dioxide from our lungs. The foods and tissues of the body are subjected to what chemical process to produce it?	Oxidation
2.776	14. What material besides ore, fuel and flux must be used in the metallurgy (smelting) of iron ore?	Hot air
2.963	8. What product besides carbon dioxide forms when magnesium carbonate is heated?	Magnesium oxide

FIGURE 2. SUMMARY SHEET

Chapter.................**Topic**...

Date....................**School**.................**Teacher**..................

Score of Questions
(How many pupils answered correctly the first question? The second? Etc.)

Question No.	Sections 1	2	3	4	5	Total	Per cent
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
Section Totals							

Score of Pupils
(How many pupils scored 0, 1, 2, 3, etc., i.e., answered no questions correctly? One? Two? Etc.)

No. of Questions Answered Correctly	Sections 1	2	3	4	5	Total	Per cent
0							
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
Total							

3.037	1. What do we call those elements whose hydroxides are bases?	Metals
3.074	12. On developing a photographic plate, by what chemical process is the metallic silver deposited in the film?	Reduction
3.112	17. If sulphuric acid and sodium bromide react, what element is likely to be liberated when the sodium sulphate and hydrobromic acid are formed?	Bromine, Br.
3.261	7. What common commercial substance is formed when sodium carbonate, quick lime and an excess of quartz are fused together?	Glass
3.376	9. Into what is calcium carbonate changed if an excess of carbon dioxide is passed into water in which the carbonate is suspended?	Calcium bicarbonate, Calcium hydrogen carbonate, Calcium acid carbonate, $Ca(HCO_3)_2$, $CaH_2(CO_3)_2$
3.492	3. Type-metal has the ability to expand on solidifying, making a fine casting. This ability of the alloy is due to what metal?	Antimony, Sb.
3.778	19. What ion is liberated in excess by hydrolysis when washing soda is dissolved in water?	Hydroxyl, OH
4.196	4. What is the approximate weight of 11.2 liters of carbon dioxide?	22 g.
4.248	16. Write the equation for the complete combustion of the third member of the methane (marsh gas) series.	$C_3H_8 + 5O_2 = 3CO_2 + 4H_2O$
4.248	18. What active parts of acetic acid are indicated by writing the formula of this acid as $HC_2H_3O_2$ instead of $H_4C_2O_2$?	Ions
4.742	20. Write a reaction for the reduction of sulphuric acid to hydrosulphuric acid by a binary acid which is a strong reducing agent?	$H_2SO_4 + 8HI = H_2S + 4H_2O + 4I_2$.

Results

These tests were given to the pupils, boys and girls, of the academic and technical high schools, who had just completed the ground covered by the tests. The results are based on the work of 581 first-semester and 268 second-semester pupils.

In this study both weighted and unweighted scores have been considered and compared, Figures 3 and 4. In weighting the scores the method of Dr. B. R. Buckingham for "scaling"[3] was used. Having determined the probable error values, the weighted values were obtained by taking the P. E. of 0.000 as equal to 3.000, and by adding the obtained P. E. values to or subtracting them from this base. The results of this weighting are given in Table I.

In comparing the difficulty of the questions in the two tests, the weighted values can be tentatively taken as they stand. It is likely that some of the questions from the first-semester test if given to the more advanced pupils would have shown slightly different values—in some instances greater, in others less. Further use of the questions will reveal the nature of these variations.

Each test was limited to the work of a single semester so that there is nothing in these tests, as given, to indicate the growth factor since they were given to different groups, one of which had received training for one semester and the other for two semesters.

A distribution of the pupils in both tests is given by schools and sections in Table II. Schools A and B require chemistry of all pupils, A in the tenth year, B in the eleventh year, while the others make it elective in the twelfth year. These twelfth-year pupils, with a negligible exception, have had this chemistry preceded by a year of physics, a condition which does not exist in Schools A and B. A score of 14 is the approximate limit of the latter, eight of their pupils (both tests considered) exceeding this score, of whom three were able to make a score of 16. Of the pupils who elected to take chemistry only five made scores of more than 18, three being perfect scores. The scores bear out the fact that the differences in ages and preparation, and the privilege of election are distinct handicaps to the younger pupils.

Figure 3 should be read as follows: School A had 242 pupils take the test; their average unweighted score was 7.68, and their average weighted score was 20.79; the average pupils per class were 27, the subject was required of all pupils during their tenth year and before the subject of physics was taken. Etc.

[3] Buckingham, B. R. *Spelling ability: its measurement and distribution.* (Teachers College, Columbia University Contributions to Education, No. 59.) New York: Teachers College, Columbia University, 1913.

Figure 4 should be read as follows: School A had 144 pupils take the test; their average unweighted score was 8.31, and their average weighted score was 20.38; the average pupils per class

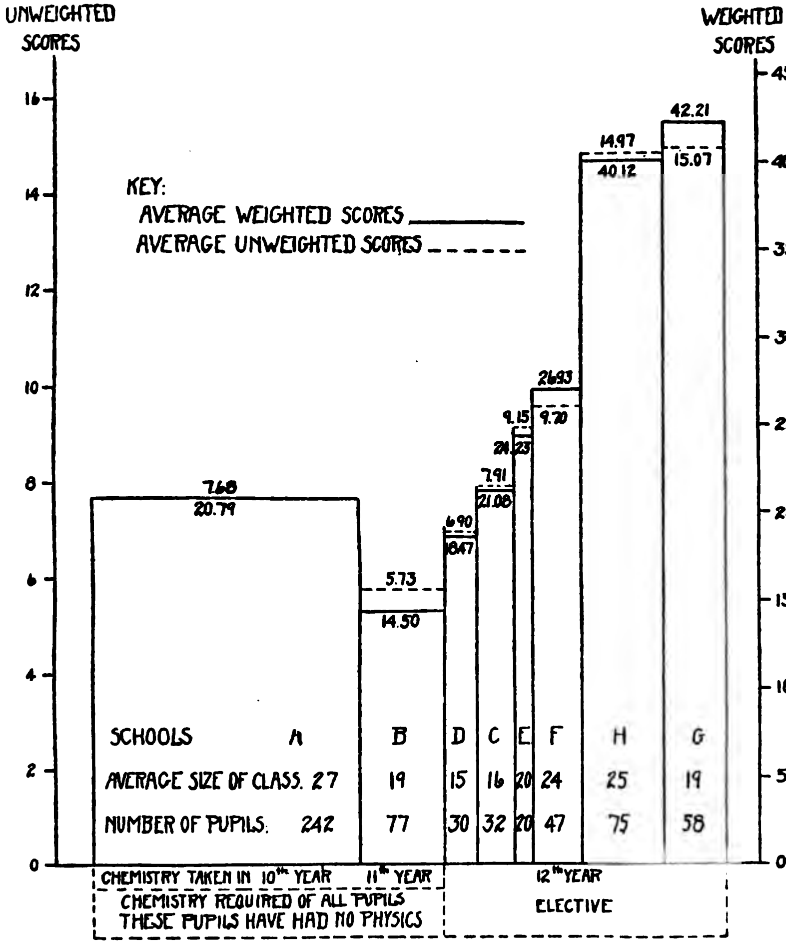


FIGURE 3. CHEMISTRY TEST, FIRST SEMESTER, JANUARY, 1920

were 21; this school required the subject of chemistry of all pupils during their tenth year and before the subject of physics was taken. Etc.

In preparing Figures 3, 4, and 5 the average score of School A, First Semester Test, was taken as a base. A comparison of the weighted and unweighted scores, Figures 3 and 4, reveals the

fact that the younger pupils of the tenth and eleventh years lost on weighting, while the older pupils of the twelfth year generally gained. The gain in chemical concepts and reasoning ability as

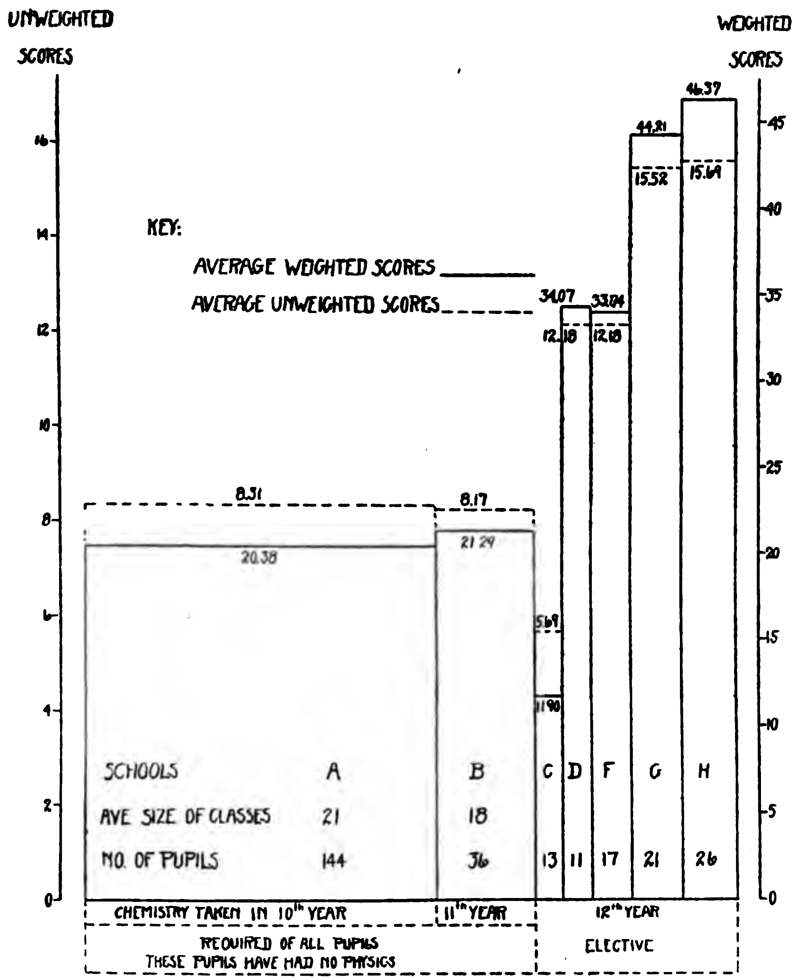


FIGURE 4. CHEMISTRY TEST, SECOND SEMESTER, JANUARY, 1920

the course proceeds is thus clearly indicated. This is more evident in Figure 5 where the averages of the two groups are compared for each semester.

Average Size of Classes

The average size of the classes which took these tests ranged from 11 to 27, while the actual size ran from 11 to 32 pupils per class. The variation in the sizes of the classes within the limits given had no consistent effect on the scores of the various classes or schools. This is according to accepted data where the classes do not generally exceed 25 pupils per class.

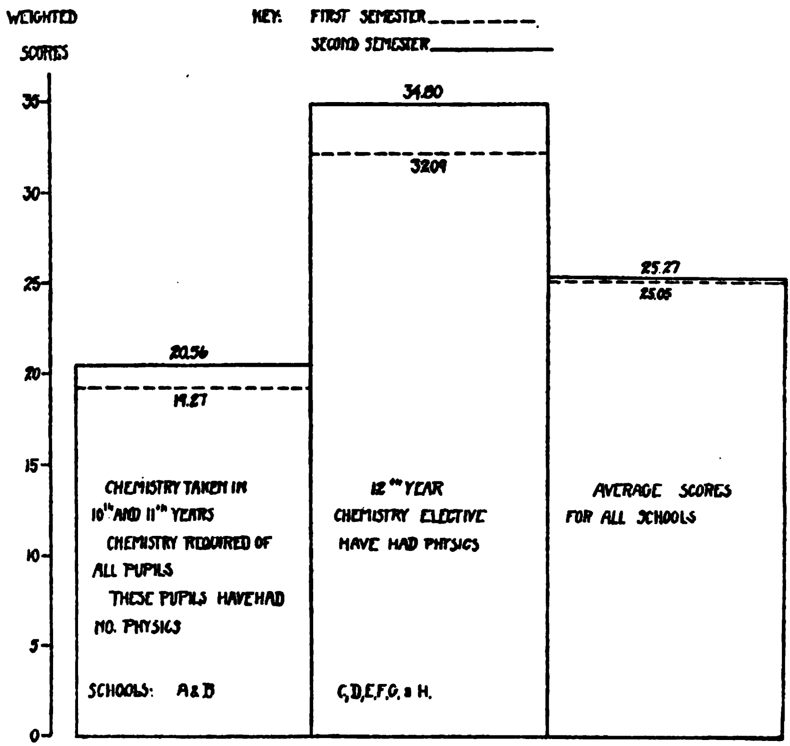


FIGURE 5. CHEMISTRY TEST, FIRST AND SECOND SEMESTERS, JANUARY, 1920—WEIGHTED SCORES

Figure 5 should be read as follows: The pupils of Schools A and B take the subject of chemistry in either the tenth or eleventh year. They had average weighted scores as follows: first semester, 19.27, second semester, 20.56. The average of all pupils taking the first semester test was 25.05, and the average of all those taking the test in the second semester was 25.27. Etc.

THE RELIABILITY OF THE BINET SCALE AND OF PEDAGOGICAL SCALES[1]

ARTHUR S. OTIS
Washington, D. C.,
AND
HERBERT E. KNOLLIN[2]
Leland Stanford Junior University

In connection with a study made by one of the writers (Knollin) on the intelligence of 180 adult males, it was suggested by Dr. Terman that it would be desirable to determine the reliability of the scale used, which was the 1915 edition of the Stanford Revision of the Binet Scale. The 180 subjects included 150 migrating unemployed and 30 business men. The usual precautions were taken in the administration of the tests.

As a measure of the reliability of the scale it was proposed to find the probable error of its score (the expression "probable error" being used in a restricted sense). This was found to be approximately six months in mental age. That is, in 50 percent of cases, mental ages of adults may be assumed to be correct within six months. It follows from this, theoretically, that in 90 percent of cases the score will probably be correct within 15 months, and in only one case in a hundred will the error probably be in excess of 23 months.[3]

[1] Involving a determination of the "probable error" of a mental age by the Binet Scale, an example of the use of a *difference* formula for correlation, and a discussion of the logical and mathematical aspects of the reliability of scales for measuring "mental" and "pedagogical" ability. This article was written in 1916. Publication was delayed on account of the war.

[2] Knollin is responsible for the testing and scoring involved in obtaining the material for this study; Otis is responsible for the method and proofs.

[3] More recently Mr. Virgil E. Dickson of Stanford University, using the same method as that described herein, has found that the probable error of a mental age when the Stanford Revision is used with first-grade children, chiefly six to eight years of age, is approximately three months. Though less in absolute amount, this is about the same proportion of the mental age. For a child of seven years, an error of three months in mental age is one of 3.57 points in intelligence quotient. Taking 14 years as the median mental age of our miscellaneous adults, an error of six months in mental age is the same error of I. Q. This indicates that the probable error of a score varies with the amount of the score, and suggests that the probable error of an I. Q. is probably approximately constant, being about 3½ points. From this it would follow theoretically that an I. Q. by the Stanford Revision is probably in error to the extent of about 6 points or more in a quarter of the cases, 10 points or more in one case in ten, and 14 points or more in one case in a hundred.

In finding the probable error of a score, a line of relation between the scores by the two halves of the scale was drawn by the "method of rank correspondence" described elsewhere.[4] The determination of a coefficient of correlation between the two series of scores by a "Difference Method" is illustrated.[5]

The method deemed proper for expressing the degree of reliability of a scale or test is to give the probable error of a score, rather than to give only the coefficient of correlation between series of scores or between degrees of difficulty of the elements of a scale for different groups of individuals.

The maker of a scale, either pedagogical or intelligence, should give, therefore, in the interest of a proper interpretation of results, the probable error of a score by the scale, as a measure of its reliability.

Theoretical Considerations with Regard to Method

There are various ways in which we may conceive of the reliability of the Binet Scale. We may ask:

1. What is the probable deviation of a mental age by the Binet Scale from the average mental age that would be obtained by the same examiner testing the same individual many times, assuming no effect remained in any case from the previous testings? This deviation would result from fluctuation of attention, etc., on the part of the examinee, and from possible differences in method of giving the tests on the part of the examiner. (*Concept 1.*)

2. What is the probable deviation of a mental age by the Binet Scale from the average mental age that would be obtained by many different examiners testing the same individual with different scales made as nearly as possible like the present Binet Scale? This deviation would result not only from the above causes but also from the differences in personality of testers, and from the impossibility of making two scales exactly alike. It would therefore be greater than the first deviation. (*Concept 2.*)

3. What would be the probable deviation of a mental age by the Binet Scale from the true mental age of the individual tested,

[4] See Reference 4 in the bibliography at the end of this article.

[5] Since this formula was first presented by one of the authors (Reference 4) he has made diligent search in the literature in order to discover whether the same formula had been presented before. Such prior presentation has not been found and the formula in its present form is believed to be original with Otis.

as determined by a hypothetical scale which measured the intelligence perfectly, according to some definition. The average of a large number of measures of the intelligence of an individual by different testers using the same and different scales would not be free from the influence of environment, etc. Accordingly, this deviation would doubtless be even greater than the second. It would be the true measure of the probable error of a mental age by the Binet Scale. (*Concept 3.*)

We have no means of measuring this true probable error since we have no true measure of the intelligence of the individual. We have no way of measuring even the second mentioned deviation since we do not have different independent Binet scales. Moreover, a second testing by a different examiner would give results somewhat affected by the first testing, and for this reason even the first-mentioned deviation must be found in an indirect way.

Still another point to be noted here is that the probable error of a single score (in the sense of being the probable deviation of the score of any test from the average of a large number of scores of the same individual by the same scale) is less than the probable deviation of any one score from another. The former is, in fact, equal to $\sqrt{1/2}$ times, or 0.707 of, the latter theoretically. (See Appendix I for proof.) It is this probable error of a single score which we are seeking ultimately and to which we shall refer when speaking of the probable error of a mental age.

For the purpose of finding the probable difference between any two measures of the mental age of an individual by the Binet Scale, made by the same tester and assuming no lasting effect of the first testing, we are obliged, since we have but one scale, to divide it into two halves and find first the probable or median difference between the mental ages of single individuals by the two halves of the scale. Upon theoretical grounds it may be shown that the probable error of a mental age by the Binet Scale as a whole is equal to $\sqrt{1/2}$ times, or 0.707 of, the probable error of a mental age by one-half the scale. Therefore the probable error of a mental age by the whole scale is equal to $\sqrt{1/2} \times \sqrt{1/2}$ times, or 1/2 of, the median difference between the mental ages by the two halves of the scale and hence can be very easily found from the latter by dividing it by 2. Proof of these propositions is given in the Appendix (I and III).

Procedure

The scale was divided into two parts by placing the first half of the tests of each age group in one scale which we have called Scale A, and the second half of the tests of each age group in a second scale which we have called Scale B. The values of each test in months were then doubled so that the mental age by each half would be comparable to that by the whole. A point was then plotted as a small circle in Figure 1 for each individual, having an abscissa equal to his score by Scale A and an ordinate equal to his score by Scale B. A relation line was then drawn through these points as shown in the figure, according to a method which is called "the method of rank correspondence," and which may be called the *single relation line* to distinguish it from a regression line of which there are two for any pair of variables. This method[6] is based on the assumption that, considering the values individually, the median of one distribution most probably corresponds to the median of the other, that the upper and lower quartile values of one distribution most probably correspond to the upper and lower quartile values respectively of the other, and similarly that the values having each of the other ranks in one distribution most probably correspond to the values having the same rank in the other distribution. For the purpose, therefore, of finding the position of the line of relation between pairs of scores, certain pairs of scores having the same ranks were selected. These were the middle values of each consecutive five in each distribution. Thus, beginning with the upper end of the distribution of Scale B values, the third, eighth, thirteenth, eighteenth, etc. values were indicated by blacking the centers of the circles; and the values having those ranks in Scale A were similarly indicated. By means of crosses points were then plotted in the quadrant whose respective abscissas and ordinates were equal to the values having the third, eighth, thirteenth, eighteenth, etc. ranks in each distribution. In this way 36 points were plotted, the abscissa and ordinate of each having the same rank in the two distributions. According to our assumption, then, these points best indicate the trend of the line of relation between the scores by the two scales. Inasmuch as there seemed to be no marked indication of a curvature in the line of relation, it was assumed to be rectilinear. Therefore a straight line was drawn as nearly as possible through the crosses, and this

[6] See also Reference 4.

was assumed to be the line of relation between the scores by the two scales.[7] If this line may be assumed to be correctly placed, then the ordinate of each point on the line represents the score by Scale B which corresponds to the score by Scale A represented by its abscissa. By means of the relation line, then, every score by Scale A may be transmuted into terms of Scale B in order to com-

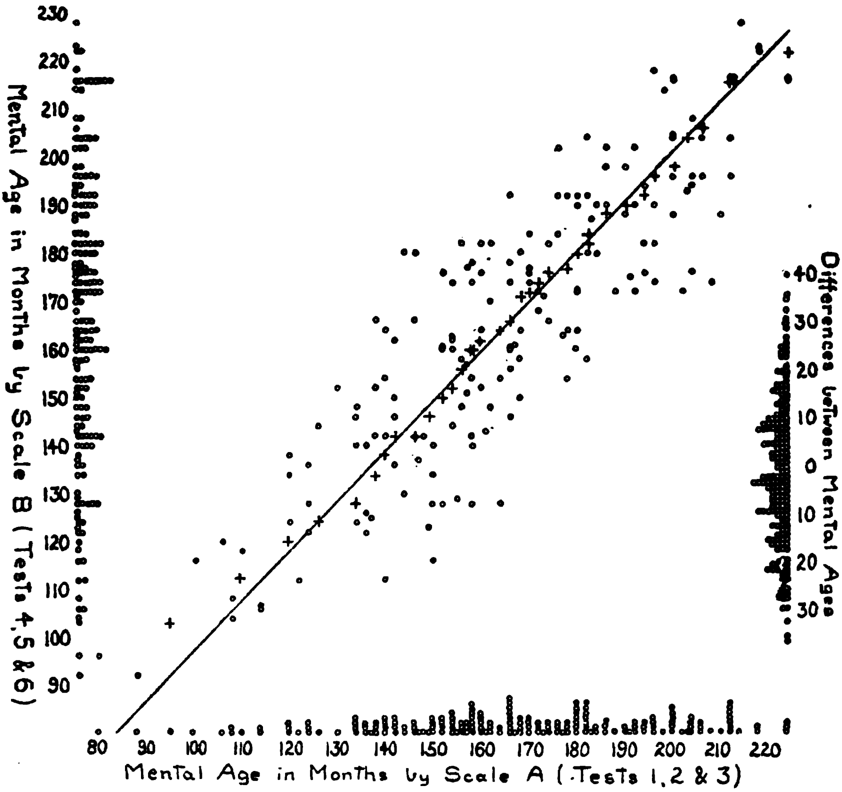


FIGURE 1. THE CORRESPONDENCE BETWEEN THE MENTAL AGES OF THE 180 INDIVIDUALS BY THE TWO HALVES OF THE BINET SCALE (STANFORD REVISION)

[7] This line is then presumably approximately as would be the line, $y = \frac{\sigma_y}{\sigma_x} x$, if the axes passed through the means of the two distributions. To draw this latter line in the first place, however, would have been to assume that the line of relation was a straight line without first determining whether it was or not. The above method makes no such *a priori* assumption.

pensate for differences in difficulty between the two scales. The difference between any actual score by Scale B and the score of the same individual by Scale A, after this has been transmuted into terms of Scale B, is one of the differences of which we are seeking the median. The value of this difference will be represented by the distance of the point for this individual above or below the line. The median of these distances, therefore, measures the probable difference, in terms of Scale B, between the scores of the several individuals in the two halves of the Binet Scale, when the difference in difficulty between them has been compensated for.

The distribution of the distances of the points above and below the line is shown in Figure 1 at the right. In order to find the median of these differences graphically with a reasonable degree of accuracy, it was necessary to construct Figure 2, in which the distribution of the differences has been plotted again at the left using a larger scale and with both plus and minus differences measured upward. These have been plotted once more at the right with each difference measured on a separate ordinate, the ordinates increasing in magnitude to the right. A smooth curve, being one-half an approximate ogive, was then drawn through these points as shown. The midpoint of this half ogive, measuring horizontally, was then found by erecting an ordinate between the ordinates of the ninetieth and ninety-first differences. This point may be seen to represent a difference of approximately 10.8 months in mental age. As a check upon this method, the average of the differences was calculated and found to be 12.62 months. Multiplying this by 0.8453[8] gave 10.67 months, which is very nearly 10.8, as the theoretical median difference. These measures, of course, are in terms of Scale B. To find the interval in terms of Scale A corresponding to 10.8 months on Scale B, it is necessary to multiply 10.8 by the cotangent of the angle of the line of relation with the horizontal axis, since this is the ratio of the projections of any section of the line of relation upon the two scales. In this case the cotangent equals 0.96. Multiplying 10.8 by 0.96 gives 10.4 months as the median difference between scores in terms of Scale A. We may now take

[8] The median deviation of a normal distribution equals 0.8453 times the average deviation.

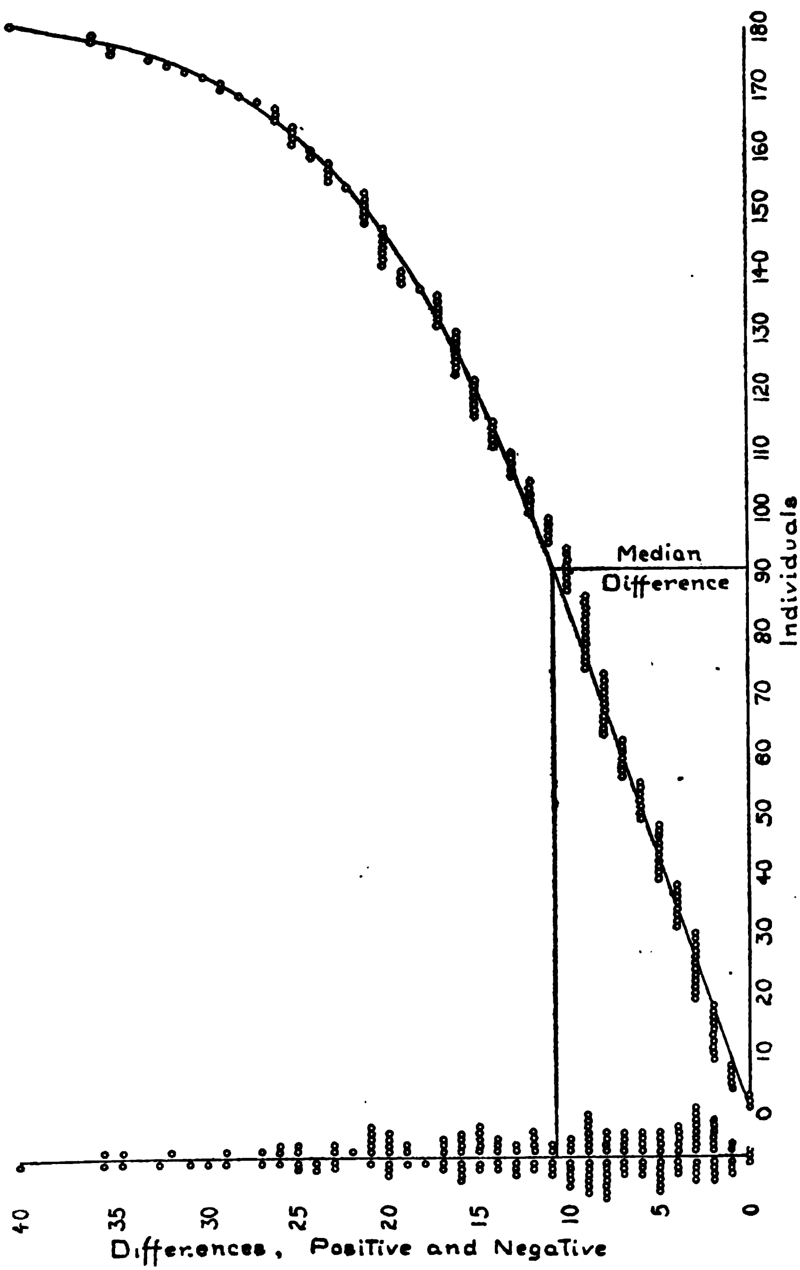


FIGURE 2. THE DERIVATION OF THE MEDIAN DEVIATION OF A DISTRIBUTION OF VALUES (DIFFERENCES SHOWN AT THE RIGHT IN FIGURE 1) BY A GRAPHIC METHOD

the average of these values, 10.8 and 10.4, which is 10.6 months, as the median difference between scores in terms of the Binet Scale as a whole. As was stated above, the probable error of the Binet Scale, according to Concept 1, is equal to one-half of this value and therefore approximately 5.3 months.

It should be noted in this connection, however, that in the cases of the highest scores—those in the neighborhood of 18 or 19 years—the differences between the scores by the two halves of the scale could not be very great since very few tests were failed. That is, these differences are less than they would have been if there had been a larger number of the more difficult tests. This, no doubt, renders our obtained value of the probable error somewhat lower than its true value. It is thought likely, therefore, that one-half year would be more nearly the true value of the probable error of an adult's score by the Stanford Revision of the Binet Scale according to Concept 1.[9]

The Improper Use of the Coefficient of Correlation as a Measure of Reliability

In this connection it might be well to make reference to a common improper use of a coefficient of correlation as an expression of the degree of reliability of a measure or scale. The manner deemed proper for expressing the degree of reliability of a score is, as in this study, to give the probable error of a score in the units in which the score is expressed. The difference between the two methods may be illustrated as follows. The correlation between the scores of the 180 individuals by the two halves of the Binet Scale has been calculated by the Pearson formula and found to be 0.850. Three other cases were considered. Calling the case of the 180 individuals Case 1, than as Case 2 were considered only the individuals whose mental ages fell between 13 and 16-11, in which case the correlation was found to be only 0.44. As Case 3 were

[9] Since writing the above, one of the writers (Knollin) has tested a number of the convicts in a California state prison, using the latest Stanford Revision of the Binet Scale (1917) and from 133 of these testings the reliability of that revision has been determined by the same method. The median difference between mental ages by half the scale was found to be 10.9 months (corresponding to 10.6 months in the case mentioned above). It may be said, therefore, that the probable error of a mental age by the later revision is practically the same for adults as that of the former and may be taken also as one-half year.

considered only those individuals whose mental ages fell between 13 and 14–11, in which case the correlation was −0.14. And as Case 4 were considered only those individuals whose mental ages were between 13 and 13–11, in which case the correlation was −0.62.

Thus it may be seen that differences in the heterogeneity of the group make very great differences in the values of the coefficients of correlation between the scores, so great, in fact, as to rob the coefficient (0.85) of much of its significance. Doubtless the correlation would have been considerably higher than 0.85 if a large number of children of ages down to 3 or 4 had been included in the group. The probable errors of the scores in the four cases were also determined and found to be respectively 5.3 months, 5.8 months, 5.9 months, and 6.0 months, showing that heterogeneity of the group, as such, probably had no serious effect upon the value of the probable error of the score. These last three values of the probable error of a score are believed to be more nearly correct than the value 5.3 for reasons stated above. The further fact that in as small a group of individuals as 31 the probable error has practically the same value as that in the cases of the larger groups speaks well for the validity of that value.

The point to be noted in this connection is that the value of a coefficient of correlation between two series of measures depends upon two variables, first, the amount of difference between the members of each pair of x- and y-values (when expressed in the same terms) and second, upon the degree of heterogeneity of the group of individuals with regard to the magnitude being measured. Now the reliability of a scale has nothing at all to do with the heterogeneity of the group, except as we wish to consider the probable error in relation to the heterogeneity. The reliability should therefore be measured by values which are independent of the heterogeneity. The probable error of a score fulfills this condition and expresses the reliability in very significant terms; it tells us what limit of error may be expected in 50 percent of cases of measurement; and from it upon theoretical grounds the limit of error that may be expected in any other percent of cases may be calculated.

It should be said that in the special case in which it is desired to find the *relative* degrees of reliability of a number of tests which have been given to the *same group* of individuals, a coefficient of

correlation between duplicate sets of scores in the same test serves as a convenient measure of the degree of reliability of that test, with which to compare a similarly derived measure of the reliability of another test. This method of comparison is valid since the degree of heterogeneity of the group is the same in all cases. The coefficients derived from two different groups of individuals, however, would not be comparable unless the degrees of heterogeneity were the same for both groups.

The Use of a Difference Formula for Correlation

It may be valuable in many cases to know the relation between the theoretical variability of the several scores of a single individual and the variability of the scores of the several individuals composing the group. A method for finding this relation has been described by one of the writers (Otis).[10] By this method a value may be obtained from a measure of the variability of differences between pairs of scores by the same individuals and a corresponding measure of the variability of the scores of the several individuals, which corresponds exactly to the Pearson coefficient of correlation between the first scores of the group of individuals and their second scores. (The second scores may be by the same scale or a similar one.) The method may be illustrated as follows.

Let us suppose, as the simplest case, that the first scores (A scores) are directly comparable to the second scores (B scores). By this is meant that differences between the A and B scores, due to practice effect or to differences in difficulty between scales, etc., may be considered negligible. Then any difference, positive or negative, between the two scores of a single individual may be considered as only chance variation. Under these circumstances presumably the variability of the B scores will be practically the same as the variability of the A scores, so that these variabilities may be considered equal for the purpose of the illustration.

Pearson has given a formula for correlation:[11]

$$r_{xy} = \frac{\sigma^2_x + \sigma^2_y - \sigma^2_v}{2\sigma_x \sigma_y} \text{ when } v = x - y$$

[10] See Reference 4.
[11] See Reference 12.

Expressed in terms of our variables, A and B, this formula becomes

$$r_{AB}=\frac{\sigma^2_A+\sigma^2_B-\sigma^2_{(A-B)}}{2\sigma_A\sigma_B}$$

Now if we assume the simplest case where $\sigma_A=\sigma_B$, we have

$$r_{AB}=\frac{2\sigma^2_A-\sigma^2_{(A-B)}}{2\sigma^2_A}\text{ or }\frac{2\sigma^2_B-\sigma^2_{(A-B)}}{2\sigma^2_B}$$

or

$$r_{AB}=1-\tfrac{1}{2}\,\frac{\sigma^2_{(A-B)}}{\sigma^2_A}\text{ or }1-\tfrac{1}{2}\,\frac{\sigma^2_{(A-B)}}{\sigma^2_B}$$

The coefficient of correlation between A and B is seen, therefore, to be the expression of a certain relation between the variability of the differences between A and B and the variability of the values of A or B themselves. That is, the coefficient of correlation increases with decrease in the ratio of these measures of variability.

If, however, the A and B scores are not directly comparable, either because of practice effect or differences in difficulty between scales A and B, then in order to find the probable error of a score by the scale, it becomes necessary, as has been explained above, to transmute the B values into terms of the A scale, or vice versa, by means of the line of relation between A and B.

In all cases of rectilinear relationship, the most probable position of the true line of relation between two variables, x and y, is the line which passes through the point in the plot representing the mean of the x values and the mean of the y values (assumed to be the origin of the plot) and which has a slope such that the tangent of the angle it makes with the horizontal axis is equal to the quotient obtained by dividing the standard deviation of the y values by the standard deviation of the x values. That is, the equation of the line which most probably expresses the true relationship between x and y is $y=\frac{\sigma_y}{\sigma_x}x$.[12] Now, if we measure the vertical deviations of the points in the plot from this line, $y=\frac{\sigma_y}{\sigma_x}x$, and call any one of these deviations d, then each value of d is the difference between a value of y and the corresponding

[12] The proof of this statement is quite involved. It is given by Otis in an article which is not yet published.

value of x after this has been transmuted into terms of y. And it may be shown that the coefficient of correlation is given by the formula,

$$r_{xy} = 1 - \frac{1}{2}\frac{\sigma_d^2}{\sigma_y^2}$$

In the previous description of this formula[18] it was called the "Deviation Formula." It has since been considered preferable, however, to call it the "Difference Formula." This formula is identical with the Pearson product-moment formula, as is demonstrated in Appendix II.

In order to express this formula in terms of our values, A and B, we must give a notation to the B scores when rendered into terms of the A scale by means of the line of relation whose equation is $A = \frac{\sigma_A}{\sigma_B} B$. Let us call these transmuted B values, B_A values. The difference formula then becomes

$$r_{AB} = 1 - \frac{1}{2}\frac{\sigma^2(A - B_A)}{\sigma^2_A} \quad \ldots\ldots\ldots\ldots\ldots \text{(a)}$$

This formula, then, expresses a certain relation between the variability of the difference between A and B values (when B values have been transmuted into terms of the A scale) and the variability of the A values. Conversely, of course, if the A values are converted into terms of the B scale, the difference formula takes the form:

$$r_{AB} = 1 - \frac{1}{2}\frac{\sigma^2(B - A_B)}{\sigma^2_B} \quad \ldots\ldots\ldots\ldots\ldots \text{(b)}$$

Two modifications of the above difference formula are

$$r_{AB} = 1 - \frac{1}{2}\left(\frac{M.\,D._{(A-B_A)}}{M.\,D._A}\right)^2 \quad \ldots\ldots\ldots \text{(c)}$$

and

$$r_{AB} = 1 - \frac{1}{2}\left(\frac{A.\,D._{(A-B_A)}}{A.\,D._A}\right)^2 \quad \ldots\ldots\ldots \text{(d)}$$

in which $M.\,D._{(A-B_A)}$ and $M.\,D._A$ are the median deviations of the distribution of $(A - B_A)$ and A values, respectively, in terms of the A scale; and $A.\,D._{(A\text{-}B_A)}$ and $A.\,D._A$ are respectively the average deviations of these distributions. As has been shown, the probable error of a score (in terms of the A scale) is expressed by the equation:

$$P.\,E._A = \frac{M.\,D._{(A-B_A)}}{\sqrt{2}}$$

[18] See Reference 4.

Hence if we wish to use the value of the probable error of a score directly in a correlation formula, we may take

$$r_{AB} = 1 - \left(\frac{P.\,E._A}{M.\,D._A}\right)^2 \ldots\ldots\ldots\ldots\ldots (e)$$

or

$$r_{AB} = 1 - \left(\frac{P.\,E._B}{M.\,D._B}\right)^2 \ldots\ldots\ldots\ldots\ldots (f)$$

Stated in words, the reliability coefficient of correlation between the duplicate scores of the individuals of a group is equal to 1 minus the square of the ratio of the probable error of a single score to the median variability of the scores, when these values are in the same terms. The advantage of enabling a correlation to be obtained directly from a value of the probable error is peculiar to the difference formula for correlation described herein. Any other corresponding measures of variability may be used, such as interquartile ranges.

To illustrate the use of form (c) of the difference formula in the present instance, the median deviation of the distribution of B values ($M.\,D._B$) was determined in the same way as the median deviation of the distribution of values (A_B-B), as shown in Figure 2. This value, $M.\,D._B$, was found to be approximately 20.1 months. Our value of $M.\,D._{(A_B\text{-}B)}$ was 10.8. Therefore, solving the formula,

$$r_{AB} = 1 - \tfrac{1}{2}\left(\frac{M.\,D._{(A_B-B)}}{M.\,D._B}\right)^2$$

we have

$$r_{AB} = 1 - \tfrac{1}{2}\left(\frac{10.8}{20.1}\right)^2$$

$$r_{AB} = 0.855$$

This is the coefficient of correlation between scores by the two halves of the scale. The corresponding value of r, found by the Pearson product-moment method, was 0.850.

The error of a coefficient of correlation found by this modification of the difference formula will depend, of course, among other things, upon the care with which the medians involved are determined, as shown in Figure 2. By the use of methods of approximation it has been found possible to obtain coefficients by the difference method about as quickly as by the method of unlike

signs, and the coefficients were believed to be much more reliable. By the use of formula (a) and (b), the reliability of coefficients is equal to that with the Pearson formula in cases of rectilinear relationship. When the relation is curvilinear the difference formula gives in certain instances a more accurate coefficient than the Pearson formula, since it corrects the coefficient, in certain instances, for skewness of the distributions in somewhat the same way as does the correlation ratio.[14] Spearman's criticism of the Pearson formula, that too great weight is given to extreme values is obviated by the use of formulas (c) or (d). These facts will partly account for slight differences between results by the two methods.

The Common Failure to Give Measures of Reliability

Improper methods of expressing reliability constitute perhaps a less prevalent fault on the part of most scale makers than the failure to give any measure whatever of the reliability of single scores. To the knowledge of the writers no measure of the probable error of a mental age by any of the several varieties of the Binet Scale has been determined before. The same is true with regard to nearly all of the pedagogical scales.

Ayres says with regard to the reliability of his spelling scale:[15] "By means of these standards children of the different grades of any locality may be tested as to their spelling attainments and the results compared with those which are found in the corresponding grades in city systems in general. . . . With less reliability the attainment of a small number of grades or of one grade may be tested. With still less reliability the attainment of a single child may be compared with these results." No measure of the P. E. of a score is given.

Thorndike says with regard to his "Scale for Measuring Ability in Reading":[16] " . . . for exact measures of individual and so for small differences amongst them, a scale with more paragraphs and questions of each degree of difficulty is required. Until such scales, Beta, Gamma, Delta, and so on, are constructed, however, the present scale is the best to use." No measure of the reliability of a score is mentioned.

[14] See Reference 11, p. 204.
[15] Reference 2, p. 40.
[16] Reference 9, 1915, p. 458.

No mention could be found of the probable error of a score in Starch's account of his grammar and arithmetic scales. He says merely, with regard to the grammar test:[17] "In spite of these limitations, which fundamentally are not of a very serious character, these tests provide measures of grammatical knowledge which are quite accurate and far more accurate than ordinary methods of testing and marking."[18]

Starch also says with regard to his reading test:[19] "It is recommended that the vocabulary test on page 37 be given in conjunction with the test for speed and comprehension. These three tests together will serve as a very adequate measure of a pupil's reading ability."

Perhaps no better comment can be made on the general attitude of makers and users of scales with regard to the reliability

[17] Reference 6, p. 626.

[18] It should be noted that no matter whether the general difficulties of the grammar tests for ten thousand children were found, the scale might still be far from reliable, due to the ambiguity of the questions and to faulty mathematics of standardization. Grammatical Scale A was submitted to three college professors of English who disagreed on the marking of sixteen different tests. Criticism could be made with regard to the theory of the construction of the scale. Such considerations, however, prove little. The proof of the pudding is in the eating and the proof of the reliability of a scale is in the P. E. of a score—at least to the extent of our saying that if the P. E. is large, the scale cannot be considered reliable. It may be invalid, as a measure of that which it purports to measure, even if the P. E. is small.

To get a rough idea of the amount of the probable error of the Starch grammar tests, Scales A and B were submitted to 25 children of grades IV to VIII of a small school. The papers were graded as accurately as possible with "keys" furnished by Starch. By comparing Scales A and B, the probable error of a score was found to be theoretically 1.2^+ steps, which, if accurate, would mean that in half the cases, a score is in error by an amount slightly greater than the difference between the abilities standard for the seventh grade and for the junior class in high school, according to standards given by Starch. To be sure, no great reliance can be placed on our figures, derived as they were from so few individuals; and it is greatly to be hoped that Starch will be able to show that they are too high. Until such time, however, the presumption is that the tests are far from being "quite accurate."

The Pearson coefficient of correlation between the scores by the two scales was found to be only 0.13, while the coefficient of correlation between the numbers of individual tests passed was 0.47. (These values are rendered comparable since the individuals were the same in both cases.) Here is evidence for the further presumption that the very simple method of counting the individual tests passed gives a more reliable score than the method advocated by Starch. No doubt the reliability of the scales could be made greater by the adoption of a better method of scoring.

[19] Reference 7, p. 21.

of the scores obtained by them than the first paragraph of chapter II of Starch's book. We will quote the paragraph with the single change of substituting the word "scales" for the word "marks." We believe it will depict with startling significance a situation which we are fast approaching. The paragraph as thus altered reads as follows:

"Scales are the universal measures of school work. Numerous and momentous problems in the operation of a school depend upon them, such as promotion, retardation, elimination, honors, eligibility for contests and societies, graduation, admission to higher institutions, recommendation for future positions and the like. Until a decade ago, no one questioned either the validity or the fairness of these measurements. It was tacitly assumed that scales were almost absolutely correct, or very nearly so, a fact attested by the surprisingly common practice of marking to a fractional part of a point, even on a hundred percentage basis."

The necessity for a measure of the reliability of the score of any test was discussed in 1912 by Otis and Davidson.[20] Obviously, no child's score in any pedagogical test can be safely used as a basis for promotion or grading, nor a test of intelligence for classification or commitment to an institution, when there are no means of taking into account the amount by which the score may be in error. Every maker of a scale for mental testing, when giving an account of the standardization of such scale, should therefore give the P. E. or some similar measure of the reliability of a score by the scale. Those who are using scales and drawing conclusions from the results are cautioned to bear in mind that every score is but an approximation, often only a very rough approximation, of the true measure of the ability in question. A mental age of fourteen years of the Binet Scale, when used with adults, is in reality a mental age of fourteen years plus or minus six months or more in half the cases. These measures of reliability are themselves subject to refinement, and at best they show only the amount of imperfection of a scale as a measure of that ability which it really measures, not necessarily the deficiency of the scale as a measure of the ability which it purports to measure. The latter deficiency may be greater.

[20] Reference 5.

APPENDIX

I. *Proof of the Formula:*

P. E. (*Single measure*) $= \sqrt{1/2} \times$ (*Median difference between measures*).

Let us suppose that we have n measures made upon a given magnitude. Let us denote these measures by m_1, m_2, m_3, etc., and assume them to be measured from the mean, M, of all the measures.

Then
$$(m_1 - m_2)^2 = m^2_1 - 2m_1m_2 + m^2_2$$
$$(m_1 - m_3)^2 = m^2_1 - 2m_1m_3 + m^2_3$$
$$(m_2 - m_3)^2 = m^2_2 - 2m_2m_3 + m^2_3$$
etc. etc.

Let us call the summation of the squares of all the differences between m's, $\Sigma(m-m)^2$. Then

$$\Sigma(m-m)^2 = (n-1)\ (m^2_1 + m^2_2 + \ldots\ldots + m^2_n) \ -m_1(m_2 + m_3 + \ldots. + m_n) \ -m_2(m_1 + m_3 + \ldots. + m_n) \ -\text{etc.}$$

Here the first term of the right member is the sum of the $n-1$ differences involving m_1, and the like number involving m_2, m_3,..........m_n. The remaining terms constitute the summation of the middle terms, $-2m_1m_2$, etc. Now to assist in simplifying this equation we will add $(m^2_1 + m^2_2 + \ldots. + m^2_n)$ to the first term and then immediately subtract these values by making the second term $-m_1(m_1 + m_2 + m_3 + \ldots. + m_n)$, etc., as shown below. Then

$$\Sigma(m-m)^2 = n(m^2_1 + m^2_2 + m^2_3 + \ldots. + m^2_n) \ -m_1(m_1 + m_2 + m_3 + \ldots\ldots + m_n) \ -m_2(m_1 + m_2 + m_3 + \ldots\ldots + m_n) \ -\text{etc.}$$

But since m_1, m_2,..........m_n are measured from their mean, their sum is zero. Hence, in the second member all terms after the first vanish. By the definition of standard deviation,

$$\sqrt{\frac{m^2_1 + m^2_2 + m^2_3 + \ldots. + m^2_n}{n}} = \sigma_m$$

or
$$m^2_1 + m^2_2 + m^2_3 + \ldots. + m^2_n = n\sigma^2_m$$
whence
$$\Sigma(m-m)^2 = n(n\sigma^2_m)$$

The first member of this equation expresses the sum of all the differences between measures. Since there are n measures, there are as many differences as there are combinations of n things taken two at a time, or $\frac{n(n-1)}{2}$ differences. Dividing by the number of differences,

$$\frac{\Sigma(m-m)^2}{n(n-1)/2} = 2\left(\frac{n}{n-1}\right)\sigma^2_m$$

Now the first member of this equation is the sum of the squares of a series of quantities, called $(m-m)$, divided by the number of such quantities. By definition, the square root of this quotient is the standard deviation of the quantities, which would be designated: $\sigma_{(m-m)}$. That is,

$$\frac{\Sigma(m-m)^2}{n(n-1)/2} = \sigma^2_{(m-m)}$$

Putting this value in place of the left member in the equation above, we have

$$\sigma^2_{(m-m)} = 2\left(\frac{n}{n-1}\right)\sigma^2_m \quad \text{from which}$$

$$\sigma^2_m = \tfrac{1}{2}\left(\frac{n-1}{n}\right)\sigma^2_{(m-m)} \quad \text{and}$$

$$\sigma_m = \sqrt{\tfrac{1}{2}}\sqrt{\frac{n-1}{n}}\ \sigma_{(m-m)}$$

Now since we have considered the measures as all having been made from the mean, we may term the values of m, observed errors, and give σ_m the notation, $\sigma_{e,\ obs.}$ Then, in this notation,

$$\sigma_{e,\ obs.} = \sqrt{\tfrac{1}{2}}\sqrt{\frac{n-1}{n}}\ \sigma_{(m-m)}$$

Now if we had an infinite number of cases, the value of $\sqrt{\frac{n-1}{n}}$ would be practically unity, and denoting the expression, "when n equals infinity" by the subscript $n=\infty$, $\sigma_{e[n=\infty]} = \sigma_{e,\ true}$, the true value of the standard error, so that $\sigma_{e,\ true} = \sqrt{\tfrac{1}{2}}\ \sigma_{(m-m)[n=\infty]}$. Now the obtained value of $\sigma_{(m-m)}$ is, of course, not always equal to exactly $\sigma_{(m-m)[n=\infty]}$. It has a probable error of $0.6745\ \frac{\sigma_{(m-m)}}{\sqrt{2n}}$. But it is the best measure we have of the true standard deviation of the differences between measures, so we may say that

$$\sigma_{e,\ true} = \sqrt{\tfrac{1}{2}}\ \sigma_{(m-m)}$$

Then since the distribution of errors and of differences between scores are approximately normal, we may assume that the median deviation of the distribution of errors, which we may call the *probable error*, is equal to 0.6745 times the standard deviation of the distribution of errors, and similarly that the median deviation of the distribution of differences $(m-m)$ is equal to 0.6745 times the standard deviation of the distribution of these differences. That is,

P. E. (*Single measure*) $= M.\,D._{e,\ true} = 0.6745\ \sigma_{e,\ true}$

and $M.\,D._{(m-m)} = 0.6745\ \sigma_{(m-m)}$

Hence we may assume that since

$$\sigma_{e,\ true} = \sqrt{\tfrac{1}{2}}\sigma_{(m-m)}$$

Therefore, *P. E.* (*Single measure*) $= \sqrt{\tfrac{1}{2}}M.\,D._{(m-m)}$

That is, the probable error of a single measure equals the-root-of-one-half times the median of the differences between the pairs of scores of single individuals. In the case of this study the single measures are scores (mental ages) by one or the other half of the Binet Scale and the differences between scores are found after inequalities of difficulty between the two halves of the scale have been compensated for.

It may be of interest to note that, following from the above proof as a corollary,

$$\sigma_{e,\ observed} = \sqrt{\frac{n-1}{n}}\ \sigma_{e,\ true}, \text{ or conversely,}$$

$$\sigma_{e,\ true} = \sqrt{\frac{n}{n-1}}\ \sigma_{e,\ observed}$$ by which formula the most

probable value of the true standard error may be derived from an observed standard error taken from any given number of measures.

II. *The Derivation of the Difference Formula and its Relation to the Pearson Product-Moment Formula*

As defined in the text, for any values of x and y, $d = y - \frac{\sigma_y}{\sigma_x}x$. For convenience, let m = the tangent of the angle of the line of relation with the horizontal axis.

Then $$m = \frac{\sigma_y}{\sigma_x}$$

Therefore, $$d = y - mx$$

Squaring, $$d^2 = y^2 - 2ymx + m^2x^2$$

Summing, $$\Sigma d^2 = \Sigma y^2 - 2m\,\Sigma xy + m^2\,\Sigma x^2$$

Transposing, $$2m\,\Sigma xy = \Sigma y^2 + m^2\,\Sigma x^2 - \Sigma d^2$$

Dividing by n, $$2m\frac{\Sigma xy}{n} = \frac{\Sigma y^2}{n} + \frac{m^2\,\Sigma x^2}{n} - \frac{\Sigma d^2}{n}$$

Or $$2m\frac{\Sigma xy}{n} = \sigma^2_y + m^2\,\sigma^2_x - \sigma^2_d$$

Dividing by $2m\sigma_x\,\sigma_y$ $$\frac{\Sigma xy}{n\sigma_x\sigma_y} = \frac{\sigma^2_y + m^2\,\sigma^2_x - \sigma^2_d}{2m\,\sigma_x\,\sigma_y}$$

Now since $$m^2\sigma^2_x = \frac{\sigma^2_y}{\sigma^2_x}\,\sigma^2_x = \sigma^2_y$$

and $$m\,\sigma_x = \frac{\sigma_y}{\sigma_x}\,\sigma_x = \sigma_y$$

therefore $$\frac{\Sigma xy}{n\,\sigma_x\,\sigma_y} = \frac{\sigma^2_y + \sigma^2_y - \sigma^2_d}{2\sigma_y\sigma_y}$$

or $$\frac{\Sigma xy}{n\,\sigma_x\,\sigma_y} = \frac{2\sigma^2_y - \sigma^2_d}{2\sigma^2_y}$$

or $$\frac{\Sigma xy}{n\,\sigma_x\,\sigma_y} = r_{xy} = 1 - \tfrac{1}{2}\,\frac{\sigma^2_d}{\sigma^2_y}$$

The left member is the Pearson product-moment formula and the right member is the difference formula.

In a normal distribution,

The median deviation (*M. D.*) of the distribution =0.6745 times the standard deviation (σ) of the distribution.

The average deviation (*A. D.*) of the distribution =0.8453 times the standard deviation of the distribution.

The interquartile range (I. Q. R.) of the distribution =1.349 times the standard deviation of the distribution.

And in cases of all distributions which are approximately normal, the same equalities are approximately true. Therefore,

$$1 - \tfrac{1}{2}\frac{(M.D.)^2_d}{(M.D.)^2_y} = 1 - \tfrac{1}{2}\left(\frac{\text{approximately } 0.6745\sigma^2_d}{\text{approximately } 0.6745\sigma^2_y}\right)$$

$$= \text{approximately } 1 - \tfrac{1}{2}\,\frac{\sigma^2_d}{\sigma^2_y}$$

Similarly,

$$1-\tfrac{1}{2}\frac{(A.D.)^2_d}{(A.D.)^2_y} = \text{approximately } 1-\tfrac{1}{2}\frac{\sigma^2_d}{\sigma^2_y}, \text{ and}$$

$$1-\tfrac{1}{2}\frac{(I.Q.R.)^2_d}{(I.Q.R.)^2_y} = \text{approximately } 1-\tfrac{1}{2}\frac{\sigma^2_d}{\sigma^2_y}$$

It may be interesting to note the relation between the difference formula and the Pearson formula:

$$r_{xy} = \frac{\sigma^2_x + \sigma^2_y - \sigma^2_v}{2\,\sigma_x\,\sigma_y} \text{ in which } v = y - x.$$

If instead of measuring the deviations (d) of the points of the plot from the line of relation, $\bar{y} = \frac{\sigma_y}{\sigma_x}x$, these deviations are measured from the line which passes through the origin making an angle of 45° with the horizontal axis, the equation of which is $y = x$, then m, the tangent of the angle of the line, $= 1$, and $d = y - x$. Beginning with this equation, the derivation of the Pearson formula is identical with the first eight steps of the derivation of the difference formula given above when all m's are omitted because in this case $m = 1$.

In a sense, therefore, this Pearson formula may be considered as a special case of the more general difference formula given above.

III. *Proof of the Theorem that the Probable Error of a Score by the Whole Scale $=\sqrt{\tfrac{1}{2}}$ times the Probable Error of a Score by Half the Scale*

We have seen from Appendix II that the reliability coefficient of correlation between the duplicate scores of single individuals is

$$r = 1 - \tfrac{1}{2}\left(\frac{\sigma_{(x-y)}}{\sigma_x}\right)^2$$

(assuming x and y to be in the same terms).

We have seen also, from Appendix I, that

$$\sigma_e = \frac{\sigma_{(x-y)}}{\sqrt{2}}$$

or

$$\sigma_{(x-y)} = \sqrt{2}\sigma_e$$

in which σ_e is the standard deviation of true errors, that is, of differences between the several measures and the corresponding true measures.

Therefore

$$r = 1 - \tfrac{1}{2}\left(\frac{\sqrt{2}\sigma_e}{\sigma_x}\right)^2$$

or

$$r = 1 - \left(\frac{\sigma_e}{\sigma_x}\right)^2 \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Now by Spearman's theorem that

$$r_{x(p),\,x(p)} = \frac{p\,r_{x(q),\,x(q)}}{q + (p-q)\,r_{x(q),\,x(q)}}$$

when p is one number of measures and q another (in this case $p = 2$ and $q = 1$),

$$r_2 = \frac{2r_1}{1 + r_1}$$

in which r_2 = the reliability coefficient of the whole scale and r_1 = the reliability coefficient of half the scale.[21]

From (1) $\quad \sigma^2_x\, r = \sigma^2_x - \sigma^2_e$

$$\sigma^2_e = \sigma^2_x - \sigma^2_x r$$

$$\sigma^2_{e_1} = \sigma^2_{x_1}(1 - r_1) \text{ half scale} \ldots\ldots\ldots\ldots (3)$$

or

$$\sigma^2_{e_2} = \sigma^2_{x_2}(1 - r_2) \text{ (whole scale)} \ldots\ldots\ldots (4)$$

(Hereafter, subscripts 1 and 2 refer respectively to the half scale and whole scale.)

Now, since the obtained score by the half scale has been multiplied by 2 (see page 123), each score by the whole scale is the average of the scores by the two half scales; that is, $x_2 = \dfrac{x_1 + x^1_1}{2}$ in which x_1 is the score by the first half of the scale and x^1_1 the score by the second half.

Whence

$$x^2_2 = \frac{x^2_1 + 2x_1x^1 + (x^1_1)^2}{4}$$

$$\Sigma x^2_2 = \frac{\Sigma x^2_1 + 2\,\Sigma^1 x_1 x^1_1 + \Sigma(x^1_1)^2}{4}$$

Since $\dfrac{\Sigma x_1 x^1_1}{n\,\sigma_{x_1}\sigma_{x^1_1}} = r_1$ and $\sigma_{x^1_1} = \sigma_{x_1}$ because the variabilities of the two half scales are presumably the same, therefore

$$\sigma^2_{x_2} = \frac{\sigma^2_{x_1} + 2r_1\sigma^2_x + \sigma^2_{x^1_1}}{4}$$

$$\sigma^2_{x_2} = \frac{\sigma^2_x(1 + r_1)}{2} \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

From (4), (5), and (2)

$$\sigma^2_{e_2} = \frac{\sigma^2_{x_1}(1 + r_1)}{2}\left(1 - \frac{2r_1}{1 + r_1}\right)$$

Simplifying, $\quad 2\sigma^2_{e_2} = \sigma^2_{x_1}(1 - r_1) \ldots\ldots\ldots\ldots\ldots\ldots (6)$

But from (3), $\quad \sigma^2_{e_1} = \sigma^2_{x_1}(1 - r_1)$

Therefore $\quad 2\sigma^2_{e_2} = \sigma^2_{e_1}$

and $\quad \sigma_{e_2} = \sqrt{1/2}\,\sigma_{e_1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$

Now since the distribution of errors is presumably normal, we may say that

P. E. (whole scale) $= 0.6745\sigma_{e_2}$

and P. E. (half scale) $= 0.6745\sigma_{e_1}$

whence P. E. (whole scale) $= \sqrt{1/2}$ P. E. (half scale)

Q. E. D.[22]

[21] Spearman, C. "Correlation calculated from faulty data," *British Journal of Psychology*, 3:271–95, October, 1910.

[22] For suggestions concerning this proof, the writers are indebted to Dr. Truman L. Kelley.

BIBLIOGRAPHY

1. Abelson, A. R. "The measurement of mental ability of 'backward' children," *British Journal of Psychology*, 4:268–314, December, 1911.
2. Ayres, Leonard P. *Measurement of ability in spelling*. New York: Russell Sage Foundation. (Educational Monograph E139.)
3. Buckingham, B. R. *Spelling ability: Its measurement and distribution*. (Teachers College Contributions to Education No. 59.) New York: Teachers College, Columbia University, 1913. 116 pp.
4. Otis, Arthur S. "Reliability of spelling scales, involving a 'deviation formula' for correlation," *School and Society*, 4: 676, 716, 750, 793, October 28–November 18, 1916.
5. Otis, Arthur S. and Davidson, P. E. "Reliability of standard scores in adding ability," *Elementary School Teacher*, 13:91-105, October, 1912.
6. Starch, Daniel. "The measurement of ability in English grammar," *Journal of Educational Psychology*, 6:615–626, December, 1915; and "A scale for measuring ability in arithmetic," *Ibid.*, 7:213–222, April, 1916.
7. Starch, Daniel. *Educational measurements*. New York: Macmillan Company, 1916. 202 pp.
8. Thorndike, Edward L. *An introduction to the theory of mental and social measurements*. (Second edition.) New York: Teachers College, Columbia University 1913. 277 pp.
9. Thorndike, Edward L. "An improved scale for measuring ability in reading," *Teachers College Record*, 16:445–467, 17:40–67, November, 1915 and January, 1916.
10. Krueger, F. und Spearman, C. "Die Korrelation zwischen verschiedenen geistigen Leistungsfähigkeiten," *Zeitschrift für Psychologie*, 44:50–114, 1906.
11. Yule, G. Udny. *An introduction to the theory of statistics*. London: C. Griffin and Company, 1919. 398 pp.
12. Pearson, Karl. "On further methods of determining correlations," Mathematical Contributions to the Theory of Evolution, XVI, Drapers Co. Research Memoirs, Biometric Series, IV.

Editorials

A HUMAN LADDER

The idea of the so-called human ladder is more or less familiar. In the army, captains in rating candidates for lieutenancies were asked to do so under several headings. One of these was "leadership." In order to rate candidates with respect to leadership, each captain was to set up his own scale. He was to evoke from his experience the best, poorest, and average lieutenants in point of leadership. Two other lieutenants—one midway between the best and the average, the other midway between the poorest and the average—completed a leadership ladder consisting of five rounds. A candidate could therefore be given the highest number of credits if the captain judged his leadership to be equal to the best within his experience. He could be given the next larger number of credits if his leadership seemed to match that of the lieutenant whom the rater had chosen to occupy the second position in his ladder, and so on.

This device has a number of defects. Because there are but five rounds, either the ladder must be short or the rounds must be far apart. If the particular captain's experience has been meager, he will probably not have encountered either the very best or the very poorest leadership. The spread of his scale (i. e., the length of his ladder) will, therefore, be small. His ratings may be relatively accurate, but only within a narrow range of the train in question. Clearly such a captain's ratings will not have the same meaning as those of a captain who has drawn upon a rich experience. In fact, the five-round human ladder is defective because each person has a different one. It can only be thoroughly satisfactory if the five standard lieutenants have been drawn from a body truly representative of all lieutenants, and if they have been objectively selected for their positions on the scale or at least selected upon the consensus of a large number of persons competent to render judgment.

Suppose it were possible to construct a ladder or scale for measuring ability in a school subject which should have one hundred steps instead of five, which should be the same for

everybody, and which should consist of steps objectively determined and objectively used. Suppose further that this scale were such that when two or more of them were constructed, each having reference to a different school subject, ratings according to each of these scales would mean the same and be capable of being combined mathematically. Would not such a device be worth while?

The percentile table possesses the properties which we have described. It exhibits a series of typical scores. In its fullest form there would be one hundred of them; but in the form commonly used there are only ten or twelve. For example, it has been shown that the eighth-grade percentile scores for Harlan's American History Test (1,691 pupils participating) run as follows:

Percentile.....	95	90	80	70	60	50	40	30	20	10	5
Score.........	95	94	92	81	75	68	57	51	41	31	28

The explanation of this table will be simplified if we assume that the children who participated were typical eighth-grade pupils. This supposition is not necessary, but if it is not warranted one must define the group from which the figures were derived.

If our group may be taken as typical, the condition we have exhibited in the above figures may be described as follows. If we had at our disposal one hundred eighth-grade children so selected that they would give thoroughly typical responses to Harlan's American History Test, and if we had these pupils arranged in order according to the size of their score on this test, and finally, if we called the pupil with the lowest score "No. 1" and the pupil with the highest score "No. 100," we should find that Pupil No. 95 had a score of 95, that Pupil No. 90 had a score of 94, that Pupil No. 80 had a score of 92, etc. It is clear that we have in the series of scores (95, 94, 92, 81, 75, etc.) a series of standards or rounds of a measuring ladder, and that we have names for these rounds or steps, namely, 95, 90, 80, 70, 60, etc. A pupil who scores 75 in this test performs like Pupil No. 60 in a group of one hundred typical eighth-grade children when they are numbered from 1 to 100 beginning with the poorest. We may, therefore, give him a score of 60 rather than of 75. If, on the other hand, his score on the test is 81, he equals the score of Pupil No. 70 in our human ladder, and we may give him a score of 70 rather than of 81. Of course, pupils in our classes will seldom receive the ratings which are actually printed in the above table. Only now and then will one earn a test score of either 75 or 81,

and only occasionally, therefore, will he be given a derived score of 60 or 70. If a child scores 76, we may interpolate to obtain his derived score. Manifestly it will be between 60 and 70. It is reasonable to place it one-sixth of the way from 60 towards 70. This would yield 62 to the nearest whole number.

The reader will note that a score of 68 on the test corresponds to that of the fiftieth or middle pupil of the arrangement. The score of 68 is, therefore, our old friend, the median. A pupil who attains this score may be given a derived score of 50.

It seems to us that the percentile method, converting, as it does, the crude scores on tests into scores in terms of rank, supplies a practical need. We received a short time ago the following query from a superintendent who has been doing a large amount of testing and considerable thinking: "Would it not be possible now in consideration of all the scores which must be ailable in some of the widely-used tests to publish more detailed standards? A median is of value for certain purposes. By it a superintendent can measure the standing of different grades and of the same grade in different buildings, but when he attempts to apply accomplishment tests as an aid to a reclassification program, the median does not give him a sufficient amount of information."

The percentile table provides precisely what this superintendent has in mind, namely, "more detailed standards." It gives the median (the 50-percentile) and it gives other standards corresponding to the tenth, twentieth, thirtieth, etc. typical pupils, as the median score corresponds to the fiftieth pupil.

Another superintendent, who was about to initiate a testing program with a view to reclassifying pupils, raised a question as to how low a score should be in one test (the scores in other tests being satisfactory) to justify a retest in the one subject in which the score was low. We do not know any better way to get at this question than on the basis of percentiles. The question involves the comparison of scores made on different tests, including probably a score in an intelligence test. We need some method of scoring which shall be the same for all the tests concerned, and the percentile method has this advantage. The superintendent of whom we have just been speaking might decide that any child required retesting who had fallen as low as the thirtieth or fortieth pupil in a representative list—in other words, as low as the 30- or 40-percentile—provided his scores in intelligence and reading

were above the 60-percentile. We do not suggest these as working bases. Any basis which seems reasonable may be chosen. Our point is that the percentile method enables us to define, so to speak, a limiting discrepancy, and to say that certain administrative action shall be taken when this limit is exceeded.

Finally, our percentile scores for different tests can be combined. The score on a spelling test can be combined with the score on an arithmetic test although the original units are entirely different. Of course, we are aware that there is a sense in which the steps of a percentile scale are not equal. For example, it will be observed that in the above table the step from the 30-percentile to the 40-percentile is six units, but that the step from the 40-percentile to the 50-percentile is eleven units. The step from the 80-percentile to the 90 percentile is only two units. Nevertheless, we assert that there is a very real sense in which these steps are equal. So far as the data are representative of typical conditions, they indicate that Pupil No. 40 is as much better than Pupil No. 30 as Pupil No. 50 is better than Pupil No. 40. They indicate that it is just as much more difficult for eighth-grade pupils to raise their score from 92 to 94 as it is to raise their score from 51 to 57. This must be so, for otherwise, the next ten pupils who exceed the performance of Pupil No. 80 could only win their distinction by obtaining larger scores. This goes pretty deeply into the question of what difficulty is. We shall not enter into it further than to say without argument that our definition of difficulty rests upon the proportion of persons who can perform the act in question. With this definition of difficulty it is perfectly justifiable to argue that it is as hard, according to our percentile distribution, for an eighth-grade child to raise his score from 92 to 94 as it is to raise his score from 57 to 68. This is because a score of 94 is so much more difficult to attain than a score of 92 that approximately 10 percent more pupils fail to get it. Likewise, 10 percent more pupils fail to reach 68 than to reach 57.

It is our judgment, therefore, that a much greater use should be made of percentile distributions than is now being made. We feel sure that this larger use of them would manifest itself, if their nature and practical utility were better understood. We recommend to research workers that in their reporting they utilize to a greater extent than they have heretofore this kind of human ladder.

B. R. B.

Reviews and Abstracts

E. H. Cameron, *Editor*

Wilson, G. M. and Hoke, Kremer J. *How to measure.* New York: Macmillan Company, 1920. 285 pp.

Continued interest in the problem of tests and measurements is reflected in a recent volume entitled *How to Measure.* The two controlling ideas of the discussions, as stated in the preface, are: "first, that the work in measurement should be handled more and more by the individual classroom teacher; and second, that the chief purpose to be served by standard tests is the diagnosis of pupil ability and pupil difficulties." This view of the purpose of tests is fundamental and cannot be over-emphasized.

The discussions in a book on measurement can be organized from at least two points of view; either the important measurable phases of each subject can be analyzed, or the tests and scales which are available can be described. The authors have chosen the latter plan and have discussed diagnosis only in so far as it can be carried on through the use of standardized tests. In this connection they do not attempt a critical evaluation of all tests. On the other hand, they discuss only those tests which on account of their use, purpose, and adaptability have been found to be most serviceable to the classroom teacher.

The purpose of the book is doubtless best reflected in chapter III entitled "The Measurement of Handwriting." In this chapter the authors have discussed in detail the abilities and classroom products which are to be measured, the methods of giving tests and scoring results, standards of attainment, and remedial instruction. If the same plan had been followed with equal thoroughness in all subjects, the book would have made a distinct contribution. In its present form, it does not excel some of the books on tests and measurements in practical value.

The book impresses the reader both favorably and unfavorably. This impression can be expressed most effectively by certain contrasts. (a) The phases of handwriting which can be measured are discussed in detail; similar discussions of other subjects appear in only a limited number of chapters. (b) Certain chapters contain the latest information concerning tests in a given subject; other chapters omit many recent developments, giving the impression that they were written some time ago. (c) The bibliography for arithmetic is well organized, fairly complete, and very suggestive; the bibliographies for certain subjects are incomplete and poorly organized. (d) The value of general intelligence tests is emphasized; recent investigations concerning the relation of age and grade standing to accomplishment in school subjects are not discussed clearly and pointedly.

A very commendable feature of the book merits special mention. It has been written in simple, clear English, which greatly increases its value to the classroom teacher.

W. S. Gray

University of Chicago

RAPEER, L. W. *The consolidated rural school.* New York: Charles Scribner's Sons, 1920. 545 pp.

This volume is a very complete discussion of the history development and problems of the consolidated rural school. The volume is based on rather definite aims of education and on a social theory of the function of the rural public school. The author states in the preface: "The general aim held is that of social efficiency while the subordinate matters under which may be grouped the principal needs of the country people and the principal problems of life which they solve are analyzed as, (1) vital efficiency; (2) vocational efficiency; (3) avocational efficiency; (4) civic efficiency; (5) moral efficiency. These are the fundamental goals of each chapter and are treated explicitly in the chapters on the program of studies." The various topics related to those problems are treated by leading specialists and successful workers in this field. The unifying idea is the organization and cooperation of rural people in meeting their life problems through the agency of the consolidated public school.

A typical consolidated school is described. Curricula for consolidated schools are set up. The advantages and disadvantages of the consolidated school are fully discussed and an extensive bibliography on the subject is listed at the close of the volume. The author seems to have gathered together about all that has been said and done on this vital problem of rural education. The book should be an inspiration and a help to rural life leaders.

A. W. NOLAN

University of Illinois

SNEDDEN, DAVID. *Vocational education.* (Brief Course Series.) New York: Macmillan Company, 1920. 500 pp.

There have been so many compilations and second-hand treatises on vocational education that it is refreshing to read a straightforward, original, authoritative discussion of the subject. Such a discussion is David Snedden's recent book. The reader may not agree with some of the conclusions, but after he has read the book, there will be no question in his mind as to what Dr. Snedden thinks.

The first three chapters, "The Meaning of Vocational Education," "The Social Need of Better Vocational Education," and "The Relation of General to Vocational Education" are a clear, unequivocal, and seemingly unassailable exposition of the principles and place of vocational work in our scheme of education. A great many of the difficulties that have impeded the progress of vocational education have been misapprehension, lack of clarity and definition in the discussions, and something of a fear lest the vocational enthusiast cherished some designs against the present school organization. These first three chapters clearly and effectively establish the essential unity of purpose in all forms of legitimate education, and especially the complementary character of the so-called general and specific forms of education.

In chapter thirteen, Dr. Snedden has done a real service in pointing out that the severest critics of vocational education have, for the most part, based their fears and their criticisms upon a misapprehension concerning the ages of the pupils for whom vocational education is designed. In the same chapter, there is a reassuring word for those who imagine a conflict between vocational work and "education for democracy."

In each of the fifteen long chapters, this book presents a discriminating and authoritative discussion of some vital phase or division of the broad field of occupational work as touched by the schools. It is a useful book. It will serve admirably as a basic text for classes in vocational or industrial education.

S. J. VAUGHN

University of Illinois

HUDSON, JAY WILLIAM. *The college and new America.* New York: D. Appleton and Company, 1920. 202 pp.

The effects of the Great War will be good and lasting if causes and ideals are now analyzed so as to give sane direction to the reconstruction of education and the civilization which it serves. Some ideals and methods fitted to a previous century's needs carry an academic respectability not consistent with their present value. A rigid and thorough-going appraisal of college and university is now due. "For many years the rebuilding of civilization will not permit anything of human worth to go to waste." Following the emergencies of war are the new and more permanent emergencies of peace. Social reconstruction calls upon the colleges because it requires skilled intelligence of a special sort, and lays upon them a new and far-reaching responsibility. "It is the immediate task of this book to define this new obligation of our American colleges to America and to the world, not only through their valuable contributions to knowledge, but through their everyday education of American youth."

The world distrusts the academic mind, sometimes ridicules it, often caricatures it, and deems it useless for the sterner purposes of life. Wide-reaching criticisms of it come with a plausibility which is almost disconcerting to the educator himself. From the implications of many of these he is defended by a careful analysis of the nature, the importance, and the peculiar demands of his work. The search for truth for its own sake is one of the professor's first obligations; we may call it his academic obligation. In the ordinary duties of life he has his everyday obligations. Beyond this is his broader obligation to the social order as a whole, the obligation to use his special knowledge to its utmost to solve the more pressing concrete social problems of the day, and to teach others to solve them. Science has larger responsibilities than those she owes to herself. Without recognition of this obligation to society the professor and his college tend to become at least unmoral. With respect to this obligation to the social order there is a rather general failure of the academic mind. A few teachers recognize it, but the great majority do not. When the scientist does assume world obligations he tends to do so as an academic mind, hindered by abstractions whose character as such he has forgotten—the defect which is at the bottom of nearly all the typical shortcomings of college education today. College professors educate by methods too intimately dependent upon their habits of abstraction in investigation. "There is a general absence of conscious educational ideals for the student, save among administrators."

America should be the chief educational motive of our colleges. "The aim of American education is to produce a definite American social order, in relation to a definite world-order." Education should be a training of the rational will rather than of the passive reason." "Subjects abandoned at graduation are an unqualified condemnation either of the worth of these subjects for the education of the given student,

or of the manner in which they are taught. They are significant not so much of the failure of the student as of the failure of his college." "The college curriculum should be to the life of the social order what the study of law becomes to the lawyer—the defining of his aims and of the means to attain them." Correlation of studies should be attained by special correlation courses in the junior and senior years. True education will define and strengthen desires worth while and teach the sure means of their fulfillment. A broad ethical culture should assume a new place in liberal education.

To define the meaning of America and the relation of the college to American life is a difficult task; but it is worth attempting and even a partial success will be useful. The business of the American college is twofold: to train men for America; to train men for the world-order, not in spite of responsibilities to the American order, but because of them.

To the question as to how these things may be one must answer that the ultimate hope is in the college professor himself. "But a significant fact obtrudes. The majority of college teachers do not recognize their obligations to the social order at all. Those who have already attained such a consciousness must spread the contagion to those who have it not." The present training for the Ph.D. degree does not afford adequate preparation for the college teacher. Just what is the best thing for the college teacher is one of the next imperative problems.

Such are a few of the propositions analyzed and maintained in this stimulating book, separated from the reasons by which they are supported and the details through which proposed reforms may be realized. It is a volume which should be read and discussed by college teachers and administrators, for it reveals the ineffectiveness of many presuppositions and uncriticized academic ideals and points the way to a better conception of conscious values and deliberate practices.

R. D. Carmichael

University of Illinois

Koos, Leonard V. *The junior high school.* New York: Harcourt, Brace and Howe, 1920. 179 pp.

From the report of the Committee of Ten in 1893 to the present time, many suggestions have been given by representative educational bodies and by educational writers for the reorganization of the public school system. These students have emphasized first one feature and then another in their attempt to picture the functions and purposes of the intermediate or junior high school. Koos is among the first to treat the subject from a broad and comprehensive point of view.

The author discusses the factors, history and extent of "The Movement for Reorganization." "The Peculiar Functions of the Junior High School" are shown in the retention of pupils, in economizing time; in recognizing individual differences; in exploration for guidance; in providing the beginnings of vocational education; in recognizing the nature of the child at adolescence; in providing the conditions for better teaching; in securing better scholarship; and in improving the disciplinary situation and socializing opportunities. "The Program of Studies" is discussed in terms of the types, variable and constant elements and the subjects of study. Other features of reorganization are departmentalization, promotion by subject, methods, the advisory system, the staff, the social organization, and the housing and equipment.

Koos's extensive collection of illustrative material shows that the practice of establishing junior high schools results first, from the altogether too common effect of

fashion which is without educational justification; second, because of administrative rea sons which are of course necessary in any scheme of education; third, because of the educational opportunities and possibilities which they furnish.

Continued experiments in public schools and colleges of education must be made to determine the educational significance of such questions as individual differences that arei due to sex, race, heredity, environment, and maturity. It is a common practice n scholastic circles to debate the relative importance of the academic versus the vocational subjects of study in terms of the needs of individuals but the merits of neither will appear until carefully controlled experiments have shown the facts. It is almost useless to talk of differentiated curricula until courses of study are written in sufficient detail so that there is genuine differentiation according to the interests, abilities, and ideals of the pupils of junior high-school age. This seems to be the next step that must be taken by students of the intermediate school.

University of Illinois P. E. BELTING

News Items and Communications

This department will contain news items regarding research workers and their activities. It will also serve as a clearing house for more formal communications on similar topics, preferably of not more than five hundred words. These communications will be printed over the signatures of the authors. Address all correspondence concerning this department to Walter S. Monroe, University of Illinois, Urbana, Illinois.

A Study of County School Systems of Oregon

Dr. Ayres' index number for evaluating the effic ency o State school systems has been applied to the county schoo systems in the state of Oregon by Professor L. L. Stetson and Professor John C. Almack of the University of Oregon. The study, made at the invitation of the state superintendent of public instruction, is published as a monograph by his office.

Intelligence Tests to Determine Promotion

The *American Schoolmaster* for June, 1921, contains a report of the use of the National Intelligence Test, Scale A, Form 1, in Wayne County, Michigan. This test was given to the children from rural schools who applied for the county examinations. The report states that the "chief object of the test was to give to those rating the examination papers an additional check on the boys and girls." This use of intelligence tests is new. This report is the first of its kind which has come to the attention of the writer but he is acquainted with other county superintendents who have made similar uses of intelligence tests. In fact, in one county in Illinois the regular county examinations were replaced by a group of standardized intelligence and educational tests.

Silent Reading Ability of Ninth-Grade Pupils

The *Educational News Bulletin* for May, 1921 contains a preliminary report on educational tests given in Wisconsin under the direction of the state department of education during the year 1920–1921. The conditions revealed in the case of silent reading are probably typical of conditions in other states. The following paragraph is quoted from the report because of the significance of the conditions

revealed. The statements are based upon scores obtained from the use of Monroe's Standardized Silent Reading Test II, Form 1 with first-year high-school pupils.

Of far more significance than the averages and ranges of performance is the very large per cent of the students who fail to come up to the sixth grade standard of 21 in comprehension. The returns show that in the average village school reporting, 37% of the students are below the sixth grade level in their ability to understand what they read, and that in the seven cities 27.5% of the freshmen are *below the same level*. The percents below this level vary from zero to 95% of the class. Now if a pupil enters high school with only sixth grade reading ability, one of two things is sure to happen. Either he will grow discouraged and drop out, or he will pass through the school at the expense of much more effort and time devoted to study than should be required. Each of these conditions is very unsatisfactory. In either case we have the community paying high salaries to high school teachers to teach classes in which from 27% to 37% *are either totally unable to comprehend the lessons assigned or do so with great difficulty*. Such a practice undoubtedly discourages the pupil, overburdens the teacher and violates every principle of economy.

Cooperative Testing in Michigan

The Bureau of Tests and Measurements of the University of Mighigan has attempted to learn the preferences of city superintendents with reference to plans for cooperative testing. The replies to certain of the questions asked are of general interest because they indicate the preferences of practical schoolmen relative to the use of tests. The following is quoted from bulletins issued by G. M. Whipple, Director of Bureau of Tests and Measurements.

If you think it worth while for 50 to 100 Michigan towns to carry out next fall a simple and definite program of cooperative testing, and if you wish to be one of the persons to take part in it, will you not devote a little time to a careful consideration of the following questions?

1. Shall intelligence testing form a part of this program?....................
2. Which school grades shall be included?....................................
3. How many school subjects shall be included?...........................
4. What test shall be used for each of the following, if included?
   (a) Arithmetic.................... (b) Reading......................
   (c) Spelling...................... (d) Writing......................
   (e) Intelligence...

Fifty-four persons, from forty-seven school systems, replied to the questionnaire concerning the proposed testing program. Tabulation of the replies reveals the following preferences:

1.	Of 47 replies,		46 wish to include an intelligence test.
2.	" 42	"	25 wish to include at least grades III to VIII in the program; 32 wish to include six or more of the school grades.
3.	" 31	"	29 wish to include three or more school subjects, and 24 wish to include four or more, though sometimes these four include the intelligence test.
4a.	" 31	"	22 wish to use the Courtis Arithmetic Test.
4b.	" 33	"	19 wish to use the Monroe Silent Reading Test.
4c.	" 47	"	24 wish to use the Ayres Spelling Scale.
4d.	" 29	"	19 wish to use the Ayres Handwriting Scale.
4e.	" 27	"	19 wish to use the National Intelligence Test.

The Practical Utility of the National Intelligence Tests

Last fall the writer received a letter from Mr. Osborne Williams, then director of research in the Atlanta (Georgia) public schools, with regard to plans for testing in the

Atlanta high schools. The final result of the correspondence which followed was a comparative study of the writer's Cross-out Scale[1] and the National Intelligence Tests which is of general interest.[2]

The purpose was to evaluate, in a practical way, the comparative usefulness of these two scales, as measures of general ability of first-year pupils. Four questions were asked. (1) How do the two scales compare as regards the cost of the materials? (2) How do they compare as regards convenience in use? (3) How do they compare as regards time and effort in scoring? (4) What is the comparative accuracy of the measures yielded, as indications of ability? These four questions will be taken up in order.

1. *How do the two scales compare as regards the cost of materials?*—It may be said, shortly, that the contrast is striking—almost absurd. Scale B (the form used) of the National Intelligence Tests is priced at $1.60 per 25 blanks. This does not include the manual of directions, however, which is priced at $.25. Nor does it include transportation charges. Comparisons are most conveniently made in terms of cost per 100 blanks. The cost of blanks for 100 pupils is then seen to be $6.40. If we suppose no more than one manual per 100 pupils needed (or one manual for each three teachers, if the rooms average 33 pupils in size) and if we suppose the shipping cost on the 100 blanks (or 1000 odd sheets) to be no more than 35 cents—certainly a reasonable figure—we have a total cost of $7.00 per hundred pupils.[3]

In contrast the Cross-out Scale is sold at a flat rate of $1.00 per hundred, or a penny apiece. Further, with each hundred blanks are included *four* of the combined directions sheet, score sheet, and record blank, with table of norms. In order to secure a complete outfit nothing need be added to this dollar rate except shipping cost, which is a flat 15 cents per 100 blanks if the goods are sent parcel post, and is frequently less if they are sent by express. Complete materials for 100 pupils cost $1.15. The contrast—$1.15 as compared with $7.00— is surely striking.

2. *How do the two scales compare as regards convenience of the blanks, in handling?*—The contrast is again striking. The pupil's blanks of the National Intelligence Tests consist of ten pages, each 8×11. In contrast to this "booklet" (as it is called in the manual) may be put the Cross-out blank, consisting of four pages each 6×9 inches.[4]

[1] Pressey, S. L. "A Brief Group Scale of Intelligence for Use in School Surveys." *Journal of Educational Psychology*, 11:89–100, February, 1920.

[2] The study should most appropriately be reported by Mr. Williams; he did all the drudgery of the giving and scoring of the tests, the writer supplying only the general method, and the final calculations. But Mr. Williams is at present suffering from a breakdown in health. So the writer is reporting the results; but he wishes it understood that the credit for the study belongs quite as much to Mr. Williams as to himself.

[3] And it is perhaps worth adding that if both Form A and Form B are used, as is recommended for accuracy in individual diagnosis, the total cost per hundred pupils becomes a good $14.00, even counting in the manual only once, since it includes directions for both forms.

[4] A fairer comparison is given in terms of paper for 100 items of score. The National Intelligence Tests take 920 square inches of blank to present 184 items of score (and two of the tests are of a type where there are 50 percent chances of a successful score, at that). Or, 500 square inches are required per one hundred items. In contrast, the Cross-out Scale required only 108 square inches per 100 items—and there are no items with more than a 20 percent opportunity for chance success.

It should also be noticed that while on the Cross-out folder there are practically no blank spaces, much space is thus wasted on the National Committee booklet; a total of three and one-half pages of paper area is thus lost. In addition, the entire first page is devoted to the name of the tests, the record of name, age, grade, and so on, and the space for summary of score; 90 square inches are certainly not required for this

The contrast with regard to accessory materials is quite as great. The manual of directions for use of the Cross-out Scale includes directions for giving the tests, scoring directions and score sheet, and record sheet, all on another four-page folder, each page is again 6×9; the norms appear on half of a mimeographed typewriter sheet. The manual of directions for the National Intelligence Tests consists of thirty-two pages of about this size; and in addition there is the double-faced score card, 8×11, and—for Scale A—a transparent key for Test 5. No record sheet is supplied.

3. *How do the scales compare as regards effort and time in scoring?*—The situation may again be indicated very briefly. The entire directions for scoring the Cross-out Scale are given in 145 words; directions for scoring the National Intelligence Tests, Scale B, are given in 801 words.[5] No special directions are required for any single test of the Cross-out Scale; the examination is so thoroughly systematized that the general directions require no qualifications whatever, from one test to the other. The National Intelligence Tests have special directions for each test.

The reason for these special directions appears when the separate tests are more carefully studied. The problem in each test of the Cross-out Scale is simply to strike out one word or number in each line of the test. In the National Intelligence Tests the method of indicating the answer differs from test to test; thus (on Scale B) the first test requires 22 answers in arabic numerals, tests 2 and 4 require the underlining of one out of four words, test 3 calls for the underlining of one of two words, test 5 calls for the writing of "D" or "S" between two terms. (Form A is even less systematized. The answers to the first test are arabic numerals, the answers to the second test are words, written by the children; the third test required underlining of two out of five words, the fourth test calls for writing "D" or "S" between terms, while the fifth requires the writing of 120 numbers.) Scoring cannot but require many special rules, when the examination is thus loosely organized. In fact, there is one test in each form which requires a total of five special provisions in order to define fully the scoring. And, in spite of all this elaborate definition, it is occasionally found necessary to leave certain special features to the scorer's judgment.[6]

So much with regard to the obtaining of the crude score. With the Cross-out Scale nothing more is necessary than to combine the crude scores for the four tests, in order to obtain the final total score; and this total score may be obtained without recopying any of the scores on the individual tests. In using the National Intelligence Tests two further processes are necessary, after the score on each test is obtained, (a) The copying of the crude scores on to the first page and (b) the weighting of certain

purpose. There is surely, take it all in all, much "waste of good white paper" in the booklet—as a publisher remarked to the writer. And it must be remembered that this booklet is not for permanent use. A single pupil uses this blank, *once*, for about twenty-five minutes; and then, after the blank is scored, it is thrown away.

[5] It might seem that the Cross-out directions must be inadequate. The writer can say only that some 100,000 of the Cross-out blanks have been used, and there has never been any question with regard to methods of scoring or any ambiguity found in the directions.

[6] Thus (Scale B, Test 1, Rule 2): "Answer may be credited if somewhat misplaced, provided it is clearly intended as the answer to the problem in question." Or (Scale A, Test 2, Rule 4): "If it seems clear that, by a slip, one answer has been put on the wrong line, and the next answers are all thus misplaced, give credit for the answers that are right even if misplaced." And, less objective than either of these, is Rule 2, Test 2, Scale A: "Acceptable answers are listed in the key. Apparently there is, however, always the possibility of the appearance of new correct responses. If the completed sentence is clearly correct, it may be given credit even though the answers do not appear in the key; but if the slightest doubt exists, if should be marked wrong."

of the tests. The process of copying across is a minor task; but, in handling a large number of blanks it mounts up, and it can be avoided by a little ingenuity, in many instances. The weighting requires time and trouble, involves further opportunity for errors in scoring and, where both "rights" and "wrongs" must be considered, involves additional labor in the obtaining of the crude score. Three tests in Scale B and four tests in Scale A call for further rehandling of the crude score. In three of the ten tests the number of rights is multiplied by 2; in three, the final score is "rights minus wrongs"; in one test the score is rights times 3.

4. *What is the comparative merit of the measure yielded by these two scales, as indications of ability?*—Now comes the surprising part of the comparison. Both scales were given to a total of 123 boys in the freshman class of the Atlanta Technical High School. Ratings as to general ability, using a rating scale modelled after the officers' rating scale used in the army, were also obtained from the teachers. In many instances the teachers felt that they did not know the pupils well enough to rate them. Finally the rating of the one teacher who knew each section best was chosen. A single rating on each child does not give as reliable an indication of ability as might be desired. But it will serve as a rough criterion for comparative purposes. The correlation between ratings and score on the Cross-out Scale was found to be 0.40. The correlation between ratings and National Intelligence Tests, Form B, was only 0.28.

The National Intelligence Tests cost nearly seven times as much as the Cross-out Scale. The materials are over nine times as bulky as the Cross-out materials. The National Intelligence Tests require some four or five times as much time to score. And the two scales appear about equal, as measures of general ability. The contrast is surely striking. The writer does not wish to press the point; and he hopes very much that others may be interested to make similar comparisons of other tests. The above correlations can only be considered of rough suggestive value. But suggestive they surely are. They suggest (it seems to the writer) that there is at present a general lack of appreciation, among test builders, of practical requirements in the way of expense and convenience—a tendency to sacrifice such practical requirements to considerations of formal test technique. It is the purpose of the present brief paper to suggest that convenience, and accuracy of measurement, may not be so incompatible after all. If such practical considerations are not more taken account of there is danger that "testing" will come to be looked upon by the superintendent as a luxury, and by the teacher as a burden. Instead, tests and scales should become an indispensible convenience, in school work.

S. L. Pressey
Ohio State University

The Use of Mental Tests in the Whitman School

When I went to the Whitman School as principal two years ago, the teachers who had been there a year or more gave as the reason for the poor quality of the work, the inferior mentality of the children. Between January and June, 1920, the teachers and I, working under the instruction and direction of Professor George W. Frasier of the State Normal School, Cheney, Washington, gave the Stanford Revision of the Binet Test to 126 children, about one-third of the school. Every member of the VIII-A class was tested. The teachers selected the rest from those whom they considered the best and the poorest in their classes.

Table I shows chronological age and grade of each child tested. It will be seen that 46 percent are correctly graded according to chronological age; 45.2 percent are retarded; less than 9 percent are accelerated.

TABLE I. AGE AND GRADE DISTRIBUTION

Age	School Grade								
	I	II	III	IV	V	VI	VII	VIII	Totals
6–6 to 7–6	23								23
7–6 to 8–6	8	5	2						15
8–6 to 9–6	3	5	11	1					20
9–6 to 10–6	1		3	7	1				12
10–6 to 11–6			4	4	2	3			13
11–6 to 12–6			1	4	4	1	1		11
12–6 to 13–6				1			2	2	5
13–6 to 14–6				1	2		2	8	13
14–6 to 15–6				1		1		8	10
15–6 to 16–6								2	2
16–6 to 17–6							1	1	2
Totals	35	10	21	19	9	5	6	21	126

Table II shows the same pupils distributed according to mental age and grade. A very noticeable change takes place. Whereas in Table I we had 45.2 percent retarded, here we have but 11.9 percent. In like manner we note 33⅓ percent are properly placed, and 54.8 percent are accelerated. This points out that, though our school is rated in the city as one having a large percent of retardation, we have in fact a large percent of acceleration, when mental age instead of chronological age is used as a basis. The median I. Q. of the 126 children is 82.

TABLE II. MENTAL AND AGE AND GRADE DISTRIBUTION

M. A.	School Grade								
	I	II	III	IV	V	VI	VII	VIII	Totals
3–6 to 4–6	1								1
4–6 to 5–6	11								11
5–6 to 6–6	13								13
6–6 to 7–6	9	2							11
7–6 to 8–6	1	7	8	2					18
8–6 to 9–6		1	12	14	1				28
9–6 to 10–6			1	3	4	2			10
10–6 to 11–6					3			1	4
11–6 to 12–6					1	3	3	6	13
12–6 to 13–6							1	1	2
13–6 to 14–6								4	4
14–6 to 15–6							1	6	7
15–6 to 16–6								2	3
16–6 to 17–6								1	1
Totals	35	10	21	19	9	5	6	21	126

The first, fourth, and eighth grades presented the most difficult problems. We sent home those in I-B whose mental age was less than five years, and we struggled along with the others. A course of study suited to the ability of the class, had to be arranged for the fourth grade. Eighth-grade pupils with mental age less than 13.6 and I. Q.

under 85 will soon drop out, for neither the grade school nor the high school meet[8] their needs.

These data, while not complete, show (1) that the teachers had grounds for their estimates of the low mentality of the children; (2) that many so-called retarded children are graded too high; (3) that, although the best out of 376 children were tested, only 7 were found to be superior (I. Q. 110 or above); (4) that the school needs to be reorganized along the lines of a changed curriculum and a reclassification of pupils.

Perhaps a few case histories will make clear the reasons for the survey, and suggest our difficulty in competing with districts where conditions are better. The following are some of the cases studied:

A. U. Age 15.6. M.A. 11.3. I.Q. 72.5. VIII-A. A. was born in Russia. Her mother is illiterate. A. is industrious, ambitious, and pleasant. If classified properly, she would enter VI-B next semester. She spent two years in the eighth grade and in that time never made a satisfactory mark. We felt it was useless to have her repeat again, and so sent her on, hoping that she might get something from new associations. She refused to go to an industrial school, and has enrolled for the commercial course in high school. She is unable to do high school work and will drop out.

O. A. Age 16.9. M.A. 11.11. I.Q. 74.4 VIII-A. O. is very nervous and uncontrolled. She attributes her poor school work to the fact that she "loses her head." She studies very hard and even spends two or three hours with her books, each day, at home. She cannot solve arithmetical problems that require any reasoning. She does not understand what she reads. Placed according to M.A. she would be in VI-A. She will take industrial work in high school which she will probably be able to do better than she has done grade work.

R. M. Age 13.8. M. A. 9.3. I.Q. 67.6 IV-A.
M. M. Age 8.4. M. A. 9.2. I. Q. 110 III-B.

R. and M. are brother and sister. R. is feebleminded, M. is superior. Both children should be in III-A. The parents are divorced. The father altho he comes from an apparently normal family is alcoholic and degenerate. He has never been able to support his family. The mother appears normal. R. is an habitual truant, is untruthful, and recently was taken into custody by the Juvenile Court for vicious conduct outside of school. M's school work is satisfactory and her conduct is excellent.

L. H. Age 14.8. M.A. 9.1. I.Q. 61.9. IV-A.
M. H. Age 16.6. M.A. 11.6. I.Q. 71.8. VII-B.

L. and M. are sisters. Their home conditions are wretched. We sent M. to the industrial school but she soon left to go to work as a housemaid. She has had several positions but is unable to make good. L's mental age warrants placing her in III-A. M. should enter VI-B next semester.

P. G. Age 12. M. A. 14.8. I.Q. 122.2. VII-A. This is the highest I.Q. in the school. P's school work is superior, and the teachers recognize his ability. He has an imbecile sister who has never been in any school.

In order to enable these children of inferior mentality to do work suited to them, I shall organize two opportunity rooms in which we shall not attempt to follow the regular course of study. Most of these children can learn to write and read, and can acquire the fundamentals of arithmetic, but they require more time than can be devoted to them in the regular classroom.

This plan, while undoubtedly crude, is a step in the right direction, and will probably lead to a more extensive reorganization with a more definite vocationalizing of the upper grades.

FRANCES WEISMAN

Principal of the Whitman School,
Spokane, Washington

National Association of Directors of Educational Research

(E. J. Ashbaugh, *Secretary and Editor*)

OHIO STATE UNIVERSITY ESTABLISHES A BUREAU OF EDUCATIONAL RESEARCH

State Law of 1915 authorized the establishment at the State University of a department of efficiency tests and survey. No funds were granted at the time since it was thought advisable to await the recommendation of the university administrators on this point. Since that time the matter has been given much consideration, and finally the time was believed to be ripe for the inauguration of the work.

In the selection of the head of the new bureau a careful canvass was made of the men of the entire country whose training and experience were such as to seem to make them eligible for the position. Finally Dr. B. R. Buckingham, Director of the Bureau of Educational Research at University of Illinois, was chosen. He has accepted the appointment, and will assume his new duties September 1.

Members of our Association will be delighted at the news of the signal honor conferred upon Dr. Buckingham. During the two years he was president of the association, he conducted its affairs with wisdom and energy. He is still a member of the executive committee and no member evinces a greater interest in its welfare.

Perhaps Dr. Buckingham's most signal achievement was the launching of the Journal which you are now reading. The success which it has gained, the great practical value which it has already been to the school men of the country, has been largely due to his editorial ability and effort.

Speaking as I am sure I do for the entire membership of our association, we congratulate Ohio State University on securing Dr. Buckingham to head its new bureau; we congratulate the school people of Ohio on the fact that a potent force in the solution of their problems has been placed at their disposal; and we extend to Dr. Buckingham our felicitations upon his new honor and opportunity and our best wishes for his success in the new field.

We are glad also to announce another promotion among our membership. Dr. Clifford Woody, for the past four years professor of education in the University of Washington, has been brought to the University of Michigan Bureau of Mental Tests and Measurements as the director of the Bureau of Mental Tests and Measurements.

We congratulate Michigan upon adding Dr. Woody to its corps; and we are very glad indeed to have Dr. Woody back where we may hope to see him at our meetings and profit by his counsel.

H. W. Anderson, Assistant Director, Educational Research, Detroit announces the establishment of tentative norms on the standardised test in typewriting upon which he has been working for the past two years. The tests are thoroughly practical,

easily administered and scored, and yield clear-cut values. Those interested in any way in typewriting work will do well to investigate this test.

Dr. H. T. Manuel, Director Educational Research, Colorado State Normal, Gunnison, Colo. has devised a Primary Group Test of General Ability. The test consists of four parts, the first and third being practice exercises while the second and fourth are the measuring instruments. Part two consists of oral directions, picture completion, logical relations, and story arrangement. Part four consists of arithmetical reasoning, memory, learning and classification.

The test is designed for children in grades I to III inclusive and does not require the child either to read or to write. The situations are presented entirely by pictures and oral instructions. The story arrangement is probably the most unique of item in the entire set. No standards of any kind are available, but Dr. Manuel would appreciate cooperation in standardization and also constructive criticism of the test.

Dean W. F. Russell of the State University of Iowa, an honorary member of our association, was appointed on a commission to advise the Chinese government concerning the establishment of a national system of schools and sailed for the orient in August. He expects to be back at the university by the opening of the second semester in February.

Ye department editor enjoyed a most pleasant summer teaching school administration in Ohio State University.

What are you doing now at the beginning of the new year?

Journal of Educational Research

OFFICIAL ORGAN OF THE NATIONAL ASSOCIATION OF DIRECTORS OF EDUCATIONAL RESEARCH

VOL. IV OCTOBER, 1921 No. 3

CONTENTS

Entered as second-class matter at the Post Office at Menasha, Wisconsin, under the act of March 3, 1879.

Journal of Educational Research

Official Organ of the National Association of Directors of Educational Research

Editor-in-Chief: **B. R. Buckingham**
Director of Bureau of Educational Research, Ohio State University, Columbus, Ohio

PUBLISHERS' NOTICE

THE JOURNAL OF EDUCATIONAL RESEARCH is edited by Doctor B. R. Buckingham and a representative staff of associate and contributing editors. It is published by the Public School Publishing Company and is the official organ of the National Association of Directors of Educational Research. In the fullest sense it belongs to the subscribers and to those who contribute to its pages.

The agreement to publish a magazine for the Bureau of Educational Research of the University of Illinois has been annulled in accordance with its terms. In fulfillment of our obligations, assumed in that agreement to our subscribers, we now furnish this Journal published under the same editor and with a representative editorial staff.

But in establishing a Journal independent of any single institution, we wish to acknowledge a debt which we feel educational progress owes to the University of Illinois for its early recognition of the efforts which so many workers in educational research were ready to make in support of a journal in this field. The name of the magazine may be changed with the next volume, beginning with the January number, in order to give expression to its representative character. If our subscribers have a preference for the name, the publishers and the editor will be very glad to receive an expression of their desires.

JOURNAL *of* EDUCATIONAL RESEARCH

VOLUME IV OCTOBER, 1921 NUMBER 3

MEASURING THE PROGRESS OF PUPILS BY MEANS OF STANDARDIZED TESTS[1]

SAMUEL S. BROOKS

District Superintendent, Winchester, New Hampshire

From time out of mind the estimate of a pupil's progress in his school work has been left to the more or less excellent judgment of his teacher, a judgment often warped by personal prejudice due to his behavior in school, his personal appearance, or his father's standing in the community. The fact that the teacher gave tests and ranked the child on the quality of his reactions to them does not necessitate a modification of the above statement. For those tests were based solely on what she *judged* the child ought to know concerning the various school subjects as a result of her particular line of instruction. She had no way of *knowing* definitely what a child of his age and grade really ought to know in order to be as well informed as other children of his age and grade in other schools. Even the grading of the papers, after they were corrected, was mostly a matter of judgment, as has been previously shown.

Some of the more unthinking teachers took the testing and grading very seriously, marked the papers very carefully on a percentage basis, and then "passed" the pupil or "flunked" him according to whether his mark was 70 or only 69. Others, realizing more or less vaguely the injustice of such a procedure, graded the pupils' work as excellent, good, fair, poor, or very poor, which they could probably have done just as accurately without giving any tests for grading purposes at all.

But there is no longer any valid excuse for such haphazard methods of measuring the results of teaching in elementary schools. The standardized tests and scales furnish us with definite

[1] This is the fifth article by Superintendent Brooks on the general topic "Putting Standardized Tests to Practical Use in Rural Schools."

norms of achievement by means of which we can compare any child's work with the median or average for his age or grade and decide justly as to whether or not he is making normal progress.

One of my purposes in using tests has been to measure the progress of pupils in their studies. Thus far we have given the tests four times in all the schools of the district. They have been given at intervals of several months so as to permit progress between tests to show plainly in the graphs. All the data from these several tests are graphically recorded and on file. The records are very interesting and highly satisfactory so far as proof of the efficiency of this method of measurement of progress is concerned although, of course, they do not always show satisfactory progress on the part of the pupils.

As heretofore stated, our plan is to give standardized tests in as many of the elementary-school subjects as possible to all the pupils in the district three times a year. They were given first in September, 1919 for grading purposes and to get a starting point from which to measure progress. In February, 1920 the tests were given again in order to find out how the pupils were progressing and particularly to discover along what lines, if any, unsatisfactory progress was being made, so that the teachers might see where increased effort or change of method was needed. In June, 1920 they were given a third time for promotion purposes.

The scores of the individual pupils in these tests were recorded, on 4×6 cards, in the form of graphs. Each time a new test was given a new graph was drawn on each pupil's card in a different color, so that at the end of June I had, for each pupil in the district above the first grade, a graph card which showed at a glance his standing in all the subjects tested for three different periods in the school year. Each teacher had duplicate cards for the pupils of her particular school.

Since my last article was written, I have devised and had printed a 5×8 graph card which is considerably more convenient than the makeshift in use last year. The graphs reproduced in this article are shown on the new form. This new card contains not only the names of the tests but also the standard scores for each of them. Directly below the name of each test is a vertical line upon which the standard scores for that test are printed at the intersections of the vertical line with the horizontal grade

lines. For instance, the sixth-grade standard score for comprehension in Monroe's Silent Reading Test is 21. Accordingly, this number is printed at the intersection of the sixth-grade line with the vertical line below "Comprehension" and under the name of that test. The fourth-grade standard score for Woody's Division Scale is 5. The figure 5 is therefore printed at the intersection of the fourth-grade line with the vertical line directly beneath "D" under "Arithmetic-Woody." Since in the Ayres Spelling Scale and in the Hahn-Lackey Geography Scale the standard scores for any particular grade vary with the column used for testing, no scores could be printed for these tests. So, merely for convenience, the Roman numerals marking the grade lines were repeated at their intersections with the verticals for these two tests. The lowest score on any test line shows the lowest grade in which that test is given. For example, Woody's Division Scale is not given below the third grade. Hence, the lowest score for this test (3) is on the third-grade line. Similarly, Starch's History Test is not given below the sixth grade.

Figures 1, 2, and 3 are copies of the graph records of three different children for the school year 1919–1920. All three were taught by the same teacher throughout the year. The graphs are given with explanations and comments for the purpose of showing a method of recording results so as to indicate at a glance how the pupils were progressing in their school work and when they were ready for promotion.

Figure 1 shows the record of an eleven-year-old girl of about average mentality. Her mental age (M.A.) was 11 years, 7 months, and her intelligence quotient (I.Q.) was 105. Hence she is a little above the average in intelligence. Her graph, resulting from the September tests and represented by the dotted line in Figure 1, falls about equally above and below the fourth-grade line. That is, she averaged about fourth-grade (end of year) ability in the tested subjects at the beginning of the school year. Hence she was placed in the class that was beginning fifth-grade work, namely, the fifth grade according to the plan discussed in my third article. The dashed line represents the scores of the the same child from the February tests and the solid line those from the June tests. The progress of the child in her studies is shown by the steady movement of the graph from below upward.

Only two subjects show little or no increase and those will be explained a little farther on.

Let us consider separately the progress made by this pupil in each subject beginning with reading. I depend mainly on Monroe's test for measuring silent reading ability. It is well standardized, perfectly objective, eliminates the memory factor, and is, to my mind, best fitted for my particular scheme. The pupil's score for rate of silent reading in September was 80. The first point, therefore, on the September curve was plotted at the intersection of the fourth-grade line with the test line, 80 being the fourth-grade standard score as shown on the card. Her score for comprehension was 17, which is halfway between the standard scores for the fourth and fifth grades. Hence, the second point on the September graph is located halfway between the fourth- and fifth-grade lines. Now note the space between the two points just located and the corresponding points on the dashed curve. This space shows the progress made by the pupil in silent reading during the first half of the school year in relation to normal annual progress represented by the distance between the two grade lines. The advance in rate of reading is particularly marked, covering as it does the space of a grade and a quarter in a half year. The advance made in comprehension is normal; that is, a half grade of progress in a half year of work.

As shown by the corresponding points on the solid-line curve, the pupil's rate of reading increased very little during the last half of the year, while progress in ability to comprehend what was read continued to be normal. The rapid increase in rate of reading was undoubtedly due to the special emphasis placed on efficient silent reading drill which was inaugurated in the Fall term and continued throughout the year. There had never before been any such drill in any of the schools. For the year, this child's progress was a grade and a half or 50 percent above normal in rate of reading and just a grade, or normal, in comprehension.

On the addition line, note the drop of the February curve below the one for September. There might be several reasons for this, the most plausible being that the child was tired or not feeling well at the time that particular test was given in February. This surmise is supported by the fact that she "came back" strong in the June tests and showed a half grade of progress for the year in addition ability.

Little progress was shown in subtraction ability; none at all for the first half of the year. But you will note that she was already up to fifth grade in both subtraction and addition at the beginning of the year. When a child's graph shows that he is well up to or above grade in any subject, the time and effort of that child is diverted to some subject in which he is below grade. One of the chief values of the tests is their diagnostic value in showing up the weak and strong places in the work of pupils or classes so that the teacher and superintendent may know where their efforts should be concentrated in order to bring about results as nearly uniform as possible. The tendency of the graphs to flatten out and more nearly approximate a straight line toward the end of the year is the direct result of this policy of placing the emphasis where it is most needed, the places where it is most needed being indicated by the earlier graphs. The ideal curve would, of course, be a straight line, denoting ability exactly equal to the grade norms in all subjects. And an ideal year's record for a fifth-grade pupil would be three straight lines the

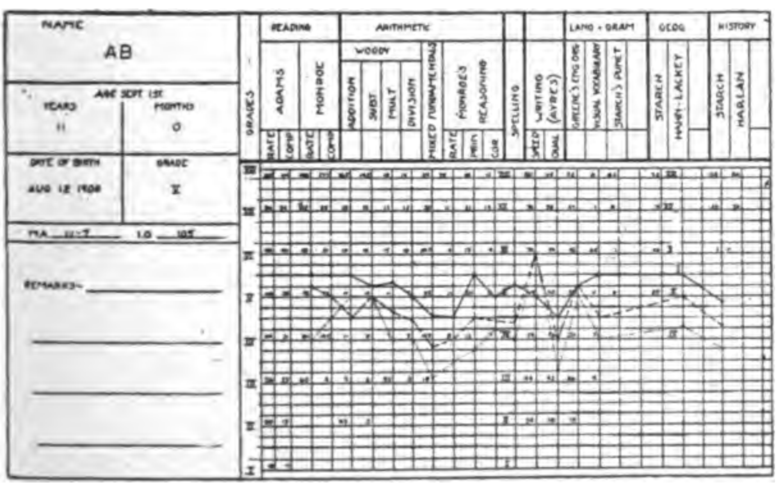


FIGURE 1. RECORD OF AN ELEVEN-YEAR-OLD GIRL OF AVERAGE ABILITY. THE DOTTED LINE REPRESENTS SEPTEMBER SCORES; THE DASHED LINE, FEBRUARY SCORES; AND THE SOLID LINE, JUNE SCORES

first coincident with the fourth-grade line on the card, the third coincident with the fifth-grade line, and the second midway between and parallel to the others. Such a record would denote absolutely even and normal progress for the year.

One of the tests given in the fall was the Cleveland Survey Test in the fundamental operations of arithmetic—a test which is excellent for purposes of diagnosis. This test showed this particular pupil to be especially weak in the multiplication and division of fractions, decimals, and denominate numbers. Special corrective drill on these phases of arithmetic was responsible for the splendid progress shown on the multiplication and division lines.

Note the very low score made in the mixed fundamentals test in September, the excellent progress made during the year, and the fact that in spite of such progress the pupil failed to come up to grade at the end of the year. It is noteworthy that only eight pupils in the whole district have so far succeeded in getting as high grades in this test as they averaged in the four fundamental operations, although the test is made up of a mixture of the identical examples used in the addition, subtraction, multiplication, and division tests. Most of them fall below from half a grade to a whole grade. A study of Figures 2, 3, and 4 reveals the same facts concerning the results from this test. Although good progress is made in every case, the pupil or class persistently grades lower in this test than in the others on fundamentals of arithmetic. To my mind this indicates that the standard scores for this test are too high.

Continuing the examination of Figure 1, we find Monroe's Reasoning Test in Arithmetic to be the next in order. This test is scored for three things: rate of solving problems, solutions correct in principle, and correct answers. Good progress is shown for the year in all three although the pupil fails to reach the grade standard for speed in solving problems.

In spelling ability the pupil accomplished 25 percent more than a normal year's progress, with nearly four times as great progress made in the last half of the year as in the first half. And here is a chance for some more interesting comparisons of the graphs on the different cards. Figure 2 shows no progress in spelling in the first half; Figure 3 shows the same; while Figure 4, which is the record of a whole fifth grade, shows considerably more progress in

the last half than in the first. The midyear tests revealed the fact that spelling work in general was progressing unsatisfactorily. As remedial measures, oral spelling drill together with Buckwalter's *Comprehensive Speller* were thrown into the discard. Ayres' Spelling Scale, supplemented by individual spelling lists made up of troublesome words from the pupils' own written vocabularies, was made the basis of the spelling course. A little booklet containing graded lists of 1600 "Common Blunder Words" was also used in most of the schools. Spelling lessons were shortened; new words were presented by a more psychological method; and the recitation consisted of a written lesson wherein the pupils use the words of the day's lesson in sentences or in a short composition. The efficacy of these changes in subject matter and method is strikingly evidenced by the greatly increased progress during the last half of the year.

Next comes handwriting. This pupil's scores in writing are typical of the general conditions revealed by the tests as discussed in my last article; speed scores up to or much above grade and quality scores very low. Although this pupil showed considerable progress for the year, she failed to reach the grade standard in quality of handwriting. But she did better than most of the pupils in this respect. Note that, throughout the year, her speed decreased while her quality increased. In the past, speed had been attained at the expense of quality. Now quality has been gained at the sacrifice of speed, and yet speed has not been reduced below the grade standard. Figure 2 also shows the fact that quality improved at the expense of speed. In most other cases, however, speed increased at approximately the same rate as quality so that the pupils were about as far behind in writing at the end of the year as they were at the beginning. All four of the records presented in this article show an improvement in handwriting for the year considerably above the average for the district. In general the improvement in writing ability was small. The reasons for the conditions found to exist at the beginning of the year and the general lack of progress during the year were fully discussed in the preceding article.

As for language and grammar, so far as the author is aware, no satisfactory general test or scale has been standardized. One of our greatest needs at present in carrying out a complete testing program in the elementary schools is a general language and

grammar test somewhat on the same plan as the Hahn-Lackey Geography Scale. Starch's Punctuation Scale is good for measuring ability in that particular. Charters Diagnostic Language and Grammar Tests are excellent as far as they go, and they cover pretty well the common errors in the use of the English language. But no standards were available for them last year, so that they did not fit into a scheme which required tests that have been fairly well standardized.[2] Hence we could do little in testing language and grammar ability last year. The two tests used, namely, Greene's English Organization Test and Thorndike's Visual Vocabulary Test might perhaps more properly be placed under the head of reading. The English Organization Test proved rather unsatisfactory. It does not seem to measure any definite ability. Its chief value seems to be in indicating, to some extent, a pupil's general intelligence or general reasoning ability, if there is such a thing, and even in this I have not found it to agree very well with the results of regular intelligence tests.

The vocabulary test, however, has proved very valuable, especially in interpreting silent reading scores. There is a high degree of correlation between the scores in the vocabulary test and those of comprehension in silent reading if the scores of children much below normal are thrown out. When a normal child fails in comprehension of silent reading, an examination of his vocabulary scores will often show a serious lack of word knowledge, which can be remedied by a definite plan of vocabulary building as explained in a former article. To such a policy is due the excellent progress shown by the pupil represented in Figure 1 as regards vocabulary knowledge. This progress is shown by the curves to be from fourth-grade ability in September to halfway between fifth- and sixth-grade ability in June. Notice that this is also the highest point reached in the silent reading scores. This test likewise measures the efficiency of whatever method of vocabulary building may be adopted.

Highly satisfactory in amount and uniformity was the progress in geography and history, as shown in Figures 1, 2, and 4, although for some reason the history scores persistently lagged behind those in geography.

[2] Standard scores for these tests are now available and we are using them as a part of our testing program.

As before mentioned, Figure 1 is the record of a pupil a little above the average in intelligence and her record shows on the average, a little more than a normal year of progress which is as it should be. Furthermore, her progress was in the direction of a more uniform ability in all subjects. The June curve is 35 percent shorter than the September curve as shown in Figure 5 (a), thus approaching much nearer the ideal curve. This fact exemplifies the value of corrective measures based on diagnosis by standardized tests.

These records are also used for promotion purposes. When a child's graph has moved upward over a space approximately equal to the distance between two grade lines he is ready to be promoted to the next grade. As before stated, the pupil whose record is shown in Figure 1 was started on fifth-grade work at the beginning of the school year. Her graph has moved upward, as shown by the solid-line curve, until it averages better than fifth grade. This shows that she had attained fifth-grade end-of-the-year standards in June and was ready for promotion to the sixth grade and to begin work in that grade the following September.

Figure 2 shows the record of a very bright eleven-year-old girl with a mental age of fifteen years and an I. Q. of 135. Although her graph showed an average of sixth-grade ability at the beginning of the year, it was considered wisest, because of her youth and various changes in the course of study, to have her take the regular sixth-grade work for that year and to prepare herself for double promotion by taking part of the seventh-grade work. Her chart shows a progress of from half a grade in rate of silent reading and spelling to two and a half grades in multiplication. In the June tests, as shown by the solid-line curve, she averaged halfway between seventh- and eighth-grade standards and was promoted to the eighth grade. Whatever of seventh-grade work she did not take along with the sixth-grade work, she will take up in the eighth grade, thus losing nothing of subject matter and gaining a whole year's time. Figure 5 (b) shows the relative lengths of this pupil's September and June curves when straightened out. The June curve is about three-fourths as long as the September curve.

Figure 3 gives the record of a very dull boy with a chronological age of 13 years, a mental age of 9 years 10 months, and an I. Q. of 76. Note the great irregularity of the September curve

and the general lack of progress throughout the year. Note that in many instances the scores of later tests fall below those of previous ones, and that the reading scores are much lower than the vocabulary scores indicating that poor reading may be due to lack of native ability and not to lack of word knowledge. This boy fell so far short of reaching fourth-grade standards in the June tests that he was not promoted to the fifth grade. He was already two years retarded. Question:—Did we do right in retarding this child another year? Problem:—What to do with cases of this kind in rural schools where special classes are out of the question, where manual trade schools are beyond the reach of the pupils, when promotion means placing the pupil wholly out of his depth, and when retardation means discouragement. This boy will probably never get beyond the fourth or fifth grade except through mistaken charity. Would it not be well to have some provision whereby such hopelessly retarded children could be permitted to leave school and engage in some useful and profitable work under the guidance of parents or other responsible persons, at least until society becomes sufficiently civilized to make provision at public expense for the proper training of such individuals? They would at least be saved from forming habits of failure and idleness which so many such children acquire during years of forced attendance at school after they have reached the limits of their mental capacities in acquiring knowledge from books. Figure 5 (c) shows the relative lengths of this pupil's September and June curves. It should be remembered that all three of these pupils were taught by the same teacher in the same way.

Figure 4 is the record of a fifth grade containing nine pupils. It shows that the entire grade has made normal progress or better in nearly every test. As usual, however, the class is weak in quality of handwriting. It is also slightly below grade in arithmetical reasoning, in mixed fundamentals, in spelling, and in geography. On the other hand, the class is considerably above standard in reading, in the fundamentals of arithmetic, in speed of writing, and in language and grammar. On the whole it shows that both teachers and pupils have done excellent work throughout the year. Relative lengths of September and June curves are shown in Figure 5 (d). The June curve is about 20 percent shorter than the September curve.

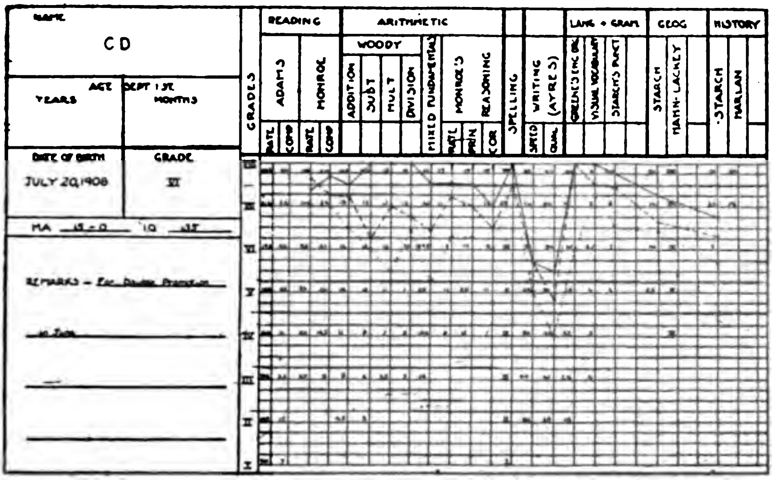


FIGURE 2. RECORD OF A BRIGHT ELEVEN-YEAR-OLD GIRL. THE DOTTED LINE REPRESENTS SEPTEMBER SCORES; THE DASHED LINE, FEBRUARY SCORES; AND THE SOLID LINE, JUNE SCORES

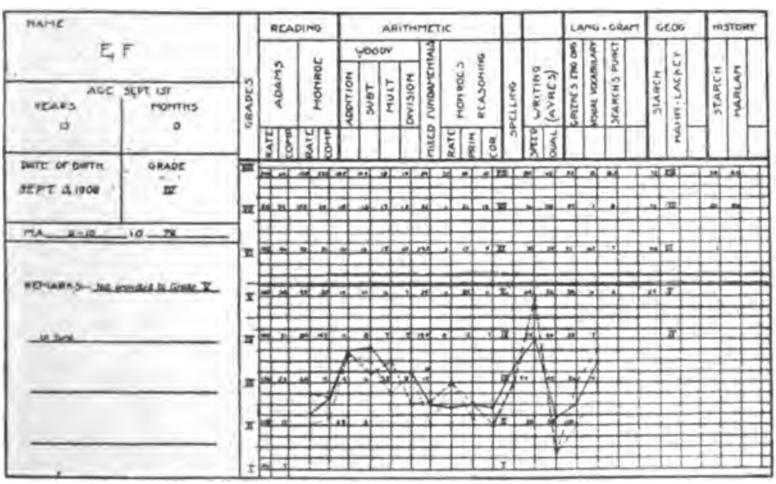


FIGURE 3. RECORD OF A VERY DULL TWELVE-YEAR-OLD BOY. THE DOTTED LINE REPRESENTS SEPTEMBER SCORES; THE DASHED LINE, FEBRUARY SCORES; AND THE SOLID LINE, JUNE SCORES

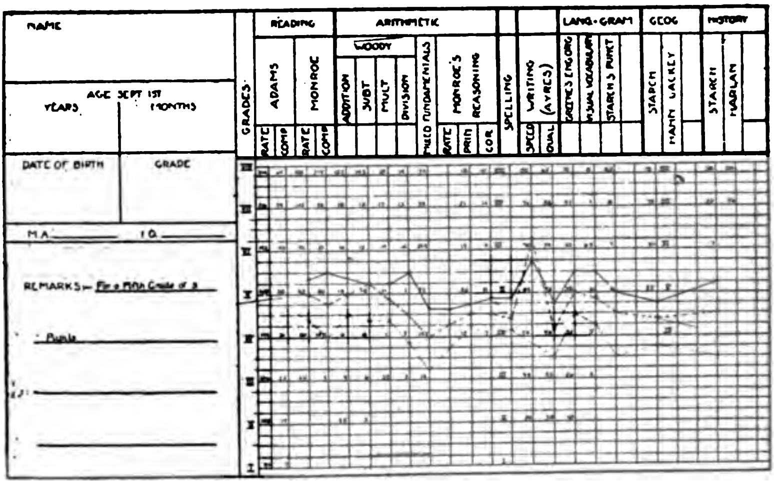


FIGURE 4. RECORD FOR A FIFTH GRADE CONSISTING OF NINE PUPILS. THE DOTTED LINE REPRESENTS SEPTEMBER SCORES; THE DASHED LINE, FEBRUARY SCORES; AND THE SOLID LINE, JUNE SCORES

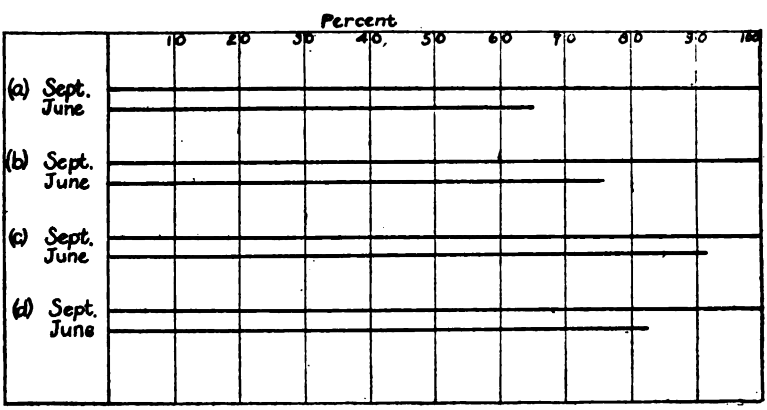


FIGURE 5. A COMPARISON OF THE LENGTHS OF SEPTEMBER AND JUNE CURVES AS SHOWN IN FIGURES 1 TO 4

VARIATION OF MARKING SYSTEMS AS DIAGNOSED BY OBJECTIVE TESTS

RIVERDA HARDING JORDAN
Dartmouth College

In recent years attention has frequently been called to the value of objective tests in diagnosing and evaluating various factors in connection with school marking, which otherwise must remain purely matters of conjecture, and the remedy for which must be applied from a purely subjective judgment. In the course of a study worked out recently, an effective concrete illustration developed which demonstrates this value very definitely. By way of bringing this use of the objective scale more especially under the observation of the school administrators, a brief description of the application made of the objective standards is here presented.

In the course of a comparison between nationality of pupils and their progress in school,[1] made in the cities of Minneapolis and St. Paul, it became necessary to collect the school marks for one semester (half a year) of the pupils in the sixth, seventh, and eighth grades. These school marks were taken from the teachers' classroom registers for ten schools in Minneapolis, involving records of 2,076 pupils. The marks were for the entire semester in each subject of instruction, and an average of all subjects for each pupil was made. In the ten schools the subjects of instruction were the same for all the pupils of any given grade, although, of course, the usual differences in the presentation of subject matter, and, indeed, in the content of subject matter offered, were found. After the average semester mark of each pupil in all his subjects was worked out, the marks were compared, and translated, where necessary, into terms of a scale of ten—i.e., a perfect mark in all subjects was 10.0, the next step in the descending scale being 9.9, and so on down to zero. Table I was then worked

[1] Jordan, R. H., Nationality and school progress, Bloomington, Illinois: Public School Publishing Co., 1921.

TABLE I. DISTRIBUTION OF MEDIAN MARKS FOR TEN MINNEAPOLIS SCHOOLS, SHOWING MEDIANS FOR BOYS AND FOR GIRLS OF EACH GRADE AND FOR EACH SCHOOL

School Number	VIIIA		VIIIB		VIIA		VIIB		VIA		VIB		All		Total
	B	G	B	G	B	G	B	G	B	G	B	G	B	G	
1[a]	7.4	8.2	7.4	8.6	6.5	8.2	6.5	8.4	6.6	6.8	7.0	7.7	6.8	8.0	7.4
2	7.8	6.8	5.5	8.4	6.5	6.6	6.4	7.0	6.7	7.8	7.8	7.3	6.8	7.3	7.1
3, 10[b]	6.0	6.7	4.7	6.7	6.6	7.6	6.3	7.0	6.2	6.5	6.3	7.7	6.0	7.0	6.5
4	5.2	6.5	5.7	6.1	5.9	6.7	6.4	6.7	5.6	6.7	7.0	7.7	6.0	6.7	6.4
5	6.2	6.7	5.1	6.4	5.2	6.4	4.8	6.5	5.5	7.6	7.5	6.4	5.7	6.7	6.2
6	5.4	7.2	4.3	6.2	5.6	7.0	4.7	6.3	4.7	7.5	3.7	7.2	4.7	6.9	5.8
7	5.7	6.7	5.4	6.7	4.8	5.8	4.7	5.3	4.9	5.3	5.2	5.3	5.1	5.9	5.5
8	5.2	6.4	4.3	6.1	4.5	5.5	5.4	6.9	5.0	3.5	5.2	5.7	4.9	5.7	5.3
9	4.5	5.7	3.0	4.5	3.0	3.1	4.4	6.7	3.7	5.3	4.8	5.2	3.9	5.1	4.5
Range..................	4.5 to 7.8	5.7 to 8.2	3.0 to 7.4	4.5 to 8.6	3.0 to 6.6	3.1 to 8.2	4.4 to 6.5	5.3 to 8.4	3.7 to 6.7	3.5 to 7.8	3.7 to 7.8	5.2 to 7.7	3.9 to 6.8	5.1 to 8.0	4.5 to 7.4

a Schools 3 and 10 are combined under one principal.
b Numbers given the schools arbitrarily to conceal their identity.

out on the basis of the medians of the boys and girls in each one of the school grades studied in each of the ten schools.

A study of the medians in the various schools, grade by grade, as well as a comparison of the school medians in the final column, brings to light a startling lack of uniformity in the marking systems employed in various buildings, although within any one building no great deviations will be noted. This condition gives rise to the feeling that in a school system of any size the old adage, "As the superintendent, so the school," must be changed to read, "As the principal, so the school."

The range of marking in the ten schools will be better understood by a study of Table II and Figure 1. It will be noted here that some of the marks run very low and some very high. As a

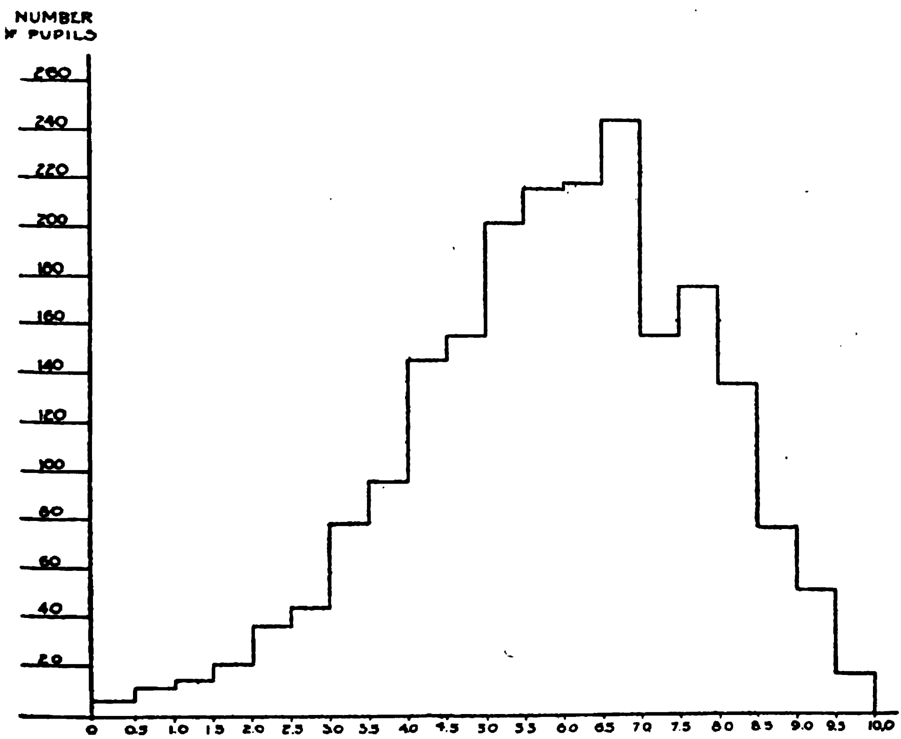


FIGURE 1. DISTRIBUTION OF SCHOOL MARKS. 10 SCHOOLS OF MINNEAPOLIS. 2,076 CASES

matter of fact, School No. 9 gives its highest ranking pupils in every grade (except the girls of VIA and VIB), a mark lower than the median pupil in the corresponding grade of School No. 1. The lower range of the scale, then, would be almost eliminated if School No. 9 were not included in the comparisons.

TABLE II. DISTRIBUTION OF AVERAGE MARKS MADE BY 2,076 PUPILS IN GRADES VIB TO VIIIA INCLUSIVE IN TEN MINNEAPOLIS SCHOOLS, IN ALL SUBJECTS

Average Mark	Boys	Girls	Total Cases
9.6–10.0	0	16	16
9.1– 9.5	7	44	51
8.6– 9.0	22	54	76
8.1– 8.5	43	91	134
7.6– 8.0	57	120	177
7.1– 7.5	61	95	156
6.6– 7.0	96	147	243
6.1– 6.5	99	116	215
5.6– 6.0	110	104	214
5.1– 5.5	107	92	199
4.6– 5.0	101	52	153
4.1– 4.5	101	45	146
3.6– 4.0	68	28	96
3.1– 3.5	55	24	79
2.6– 3.0	32	15	47
2.1– 2.5	25	10	35
1.6– 2.0	14	5	19
1.1– 1.5	10	0	10
0.6– 1.0	4	2	6
0.4– 0.5	4	0	4
Total................	1,016	1,060	2,076

In conversation with the principals of such widely varying schools as School No. 1 and School No. 8 or School No. 9, an attempt was made to determine the basis which the principal had in mind in directing the marking system of the teachers under his

charge. The principal of School No. 1 felt that the high grades given by his teachers were to be explained on a basis of the high intelligence shown by the pupils of the school, and an attempt at justification of this position was made by calling attention to the fact that the pupils came from homes of intelligence and a certain degree of wealth. The principal of School No. 9 felt that the "standard" of his school was best maintained by a very strict marking system, and was very proud of the fact that pupils, who had been given higher marks in their own buildings, were not able, when transferred from other parts of the city, to meet the requirements of his own school, and frequently had to be demoted. He felt that a low marking system was an evidence of a high standard for his building. Such responses were typical of all the schools, and it became very evident that the problem of the superintendent of bringing about a uniform marking system was almost impossible unless the principals could be convinced by some purely objective method that their theories of marking were in error.

Some months later it became possible for the writer to evaluate by objective methods the achievement of a portion of the pupils already studied, by means of certain tests designed to measure specific abilities of the children. By this time, the eighth-grade children had been promoted to the high school, so that the tests were given only to the pupils who had originally been studied in the sixth and seventh grades. The tests given were two Trabue Completion tests, two vocabulary tests, two number completion tests, two geometrical forms tests, two substitution tests, a memory span test and an opposites test. All pupils were eliminated from these tests whose marks had not been considered in the original investigation. The averages of the schools is given in Table III in terms of the raw scores for each test. For puposes of comparison these tests were worked out for schools 1, 8 and 9 only, in order to determine the relative standing of these most widely divergent marking systems.

This comparison shows a clear superiority of the pupils in School No. 8 over School No. 1, and certainly demonstrates that the contention of the principal of School No. 1, that his pupils were of higher intellectual grade than those of most of the other

schools of the city, is entirely without foundation. School No. 9 is clearly below the grade of School No. 1, but not nearly to such

TABLE III. AVERAGE SCORES FOR EACH TEST IN SCHOOL 1, 8, 9

School	Opp.	Trab.	Vocab.	Subst.	Mem. Sp.	No. Comp.	Geom. Forms
1	49.7	12.6	56.9	74.9	15.2	12.3	7.1
8	57.3	13.3	58.1	79.9	16.1	12.5	8.2
9	45.9	11.5	54.4	71.0	16.4	10.1	5.5

an extent as is indicated by the wide disparity between the marks given the pupils in the two schools. The objective tests have here solved what otherwise would be an extremely difficult situation for a school superintendent to handle. It is shown very clearly that there is no justification for an assumption on the part of any one of these principals that his subjective measurement based upon long experience is a safe or proper basis for marking his pupils.

The principal of School No. 1 will at once discover his error in assuming that his pupils are superior to the personnel of the other schools of the city. The principal of School No. 9 will realize that his low marking is due in part to a lower standard of intelligence, and therefore that it does not imply necessarily a higher standard for his school, and he will see further that an extremely low standard of marking is not justified in comparison with the other schools of the city. The principal of School No. 8 likewise will realize either that his marking scale has been unfortunate or that he is not getting the degree of attainment from his pupils which he may reasonably expect.

As a result of these tests, school principals will find it incumbent to come to some agreement among themselves as to marking standards. Not only is the result of these tests of values with reference to the general conditions as brought out by medians or by averages, but the extreme range of marking shown by the total distribution of the 2,076 pupils will be very much restricted. The curve of distribution will much more nearly approach the

normal curve when the results of the objective tests are studied and intelligently applied. The ease of application of such tests as those used, or similar intelligence tests, makes it a very simple matter for any superintendent of schools to diagnose the exact situation, and hence to evaluate properly differences in marking which exist within his school system. It is to be hoped that such diagnostic methods will be used very commonly in the immediate future.

THE RATE OF PROGRESS IN TEACHER PREPARATION[1]

W. RANDOLPH BURGESS

Formerly Russell Sage Foundation

Table I shows the improvement which has taken place in the education of teachers in service in ten states since the year 1910. The index numbers[2] used to show the standing of each state are derived from state school reports which give the number of college graduates and the number of normal-school graduates on the teaching staff. The index numbers show the number of years of college or normal-school training the average teacher has had, counting only years of study leading to graduation. In Massachusetts in 1910, for example, the average teacher had one and two-thirds years of college or normal-school training. By 1918 this record had improved; so that if the completed normal-school and college courses were evenly distributed among all the teachers there would be more than two years for each teacher.

In Table I a number of missing years have been filled in by interpolation and in the case of West Virginia the index number for 1910 was estimated. The figures have also been carefully edited for inconsistencies. The states are listed in the order of their rank in 1918.

An average index number for the ten states is given at the foot of the table. This was computed from the index numbers, giving each state equal weight, a method which gives a more stable figure than a weighted average which is largely influenced by fluctuations in one or two states with large numbers of teachers. For the years 1919 and 1920 the average is computed by the method of relatives; that is the percentage increase over the 1918 level is computed in each state where the data are available, and the average of these percents is used to determine the average index figure.

The average shows a continuous improvement amounting in the aggregate to a 42 percent increase over the 1910 figures. In the light of a recent nation-wide teacher shortage it is notable that the

[1] The second of two articles on the education of teachers in service in the United States.

[2] Burgess, W. Randolph. "The education of teachers in fourteen states," *Journal of Educational Research*, 3:161–72, March, 1921.

TABLE I. INDEX NUMBERS FOR EDUCATION OF TEACHERS IN SERVICE IN TEN STATES. 1910 TO 1920

State	Years of Completed College or Normal-School Training per Teacher										
	1910	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920
Massachusetts	1.65	1.65	1.68	1.68	1.75	1.82	1.90	1.98	2.05	2.05	2.05
Rhode Island	1.78	1.81	1.85	1.86	1.87	1.91	1.94	1.97	2.00	.....	.....
New Jersey	1.51	1.54	1.62	1.75	1.75	1.83	1.90	1.94	1.95	1.96	1.99
New Hampshire	1.00	1.09	1.17	1.23	1.30	1.36	1.43	1.51	1.59	1.65	1.71
Minnesota	0.96	1.00	1.07	1.07	1.13	1.16	1.19	1.22	1.25	1.26	1.28
Montana	0.99	1.18	1.22	1.19	1.15	1.19	1.06	1.19	1.18	1.17	1.17
Illinois	0.37	0.38	0.49	0.77	0.81	0.82	0.82	0.98	1.05	1.02	.....
Virginia	0.63	0.61	0.67	0.73	0.66	0.73	0.79	0.77	0.82	0.85	.....
Kansas	0.45	0.50	0.55	0.58	0.62	0.63	0.65	0.72	0.79	.....	.....
West Virginia	0.33	0.34	0.47	0.53	0.54	0.55	0.60	0.70	0.76	.....	.....
Average	0.97	1.01	1.08	1.14	1.16	1.20	1.23	1.30	1.34	1.35[a]	1.37
Percent increase since 1910	.....	4	11	18	20	24	27	34	39	40	42

[a] Not the averages of figures listed for 1919 and 1920, but computed from percentage increases shown by available figures in past two years.

records do not show any general falling off in the past two years. In Montana and Illinois the 1919 figures are slightly under those for 1918. In the other states there is either no loss or a slight gain; so that the average shows a slight gain. The rapid progress of previous years is interrupted, but there is no evident loss. While this showing is exceedingly encouraging it should be borne in mind that some of the effects of high prices and low teachers' salaries have not yet been felt. Decreased enrollment in normal school lessens the supply of trained teachers, not this year but in two or three years. It remains to be seen whether salary increases and better working conditions will come into play rapidly enough to offset the unfavorable influences. Apparently the situation is now better, in these states at least, than has commonly been thought.

Measuring Rates of Change

The figures of Table I are of significance, not alone for the accomplishments which they show, but also for the direction and rate of movement which they record for each state. From the

data of Table I, coefficients of regression[3] have been computed to show for each state the annual rate of progress in teacher preparation. These coefficients are shown in Table II. Illinois showed the most rapid progress during the ten-year period with an annual advance of more than seven hundredths of a year of training for each teacher. The rate is due to low figures in the first few years rather than to consistently rapid progress. It should be added that the internal evidence for the accuracy of the Illinois figures is less convincing than in the case of any other state. New Hampshire follows closely on the heels of Illinois with a remarkably steady increase in the education of its teaching staff. Montana shows the least progress, although its ranking in teacher preparation has been at or near the top of the western states for which we have figures. The state of Minnesota shows exactly the same regression coefficient as the average for the ten states.

TABLE II. COEFFICIENTS OF REGRESSION SHOWING ANNUAL RATE OF PROGRESS IN THE EDUCATION OF THE TEACHING FORCE IN TEN STATES, 1910 TO 1920

State	Annual Increase in Years of College or Normal School Training per Teacher
Illinois	0.073
New Hampshire	0.070
New Jersey	0.051
Massachusetts	0.050
West Virginia	0.048
Minnesota	0.041
Kansas	0.037
Virginia	0.025
Rhode Island	0.025
Montana	0.006
Ten states	0.041

[3] For the method employed see article by Leonard P. Ayres in *Journal of Educational Research*, May 1920, and discussion in *Trends of School Costs* by W. Randolph Burgess, Russell Sage Foundation, 1920.

Looking Forward to 1950

The coefficients of regression of Table II make possible a speculation as to the education of teachers in coming years. They furnish a method of determining where the present rate of progress will carry us. Table III shows what the education of the teachers of ten states will be in the year 1950 if the rate of progress of the past eleven years is consistently maintained.

TABLE III. INDEX NUMBERS FOR TEACHER PREPARATION IN 1950 ON THE BASIS OF THE RATE OF PROGRESS FROM 1910 TO 1920

State	Years of Completed College or Normal-School Training per Teacher
New Hampshire...	3.82
New Jersey.......	3.59
Massachusetts.....	3.58
Illinois...........	3.33
Rhode Island......	2.78
Minnesota........	2.58
West Virginia.....	2.26
Kansas...........	1.94
Virginia..........	1.61
Montana.........	1.36
Ten states..........	2.60

The figures are in terms of index numbers showing the years of completed college or normal-school training per teacher. The record of the highest state, New Hampshire, is very close to that goal so often mentioned, four years of college or normal-school training. Montana, the low state, shows a figure only a few points higher than her index in 1920. The present rate of progress in three states in the lower half of the table will not bring them in thirty years to a point where their average teacher will have two years of college or normal-school training.

Relative Standing at Three Periods

The relative ranking of the ten states at the three periods 1910, 1920, and 1950, according to the size of their index numbers

is shown in Table IV. Rhode Island drops from first place in 1910 to second in 1920 and fifth in 1950. Montana shows as great a decline in relative position, from fifth to tenth place. New Hampshire and Illinois show the greatest increases in rank.

TABLE IV. RANK OF TEN STATES IN TEACHER PREPARATION, 1910, 1920, AND 1950 (COMPUTED)

State	Rank		
	1910	1920	1950
Rhode Island	1	2	5
Massachusetts	2	1	3
New Jersey	3	3	2
New Hampshire	4	4	1
Montana	5	6	10
Minnesota	6	5	6
Virginia	7	8	9
Kansas	8	9	8
Illinois	9	7	4
West Virginia	10	10	7

Rapid progress is not confined to any one section of the country nor is slow progress characteristic of some single section. West and East divide the honors of the two states making most rapid progress and each furnishes a representative for lowest place. There is no tendency for urban states to make more rapid progress than rural. One of the two states with the lowest rates of gain is Rhode Island, the most thickly populated state in the Union, and the other is Montana, largely rural.

Another factor which might be expected to influence the rate of gain is the rate of increase in the population. Montana is a striking case in point. From 1910 to 1920 the population of Montana increased 45.6 per cent, an increase nearly twice as large as that in any other of the ten states. It is evident that with such an influx of new population the problem of finding trained teachers for rapidly growing schools is far more difficult than that of finding them in a more static population. It would seem reasonable to account for Montana's low rate of increase in teacher preparation by her large increase in population. When we exam-

ine this relationship in the other states, however, we find no clear-cut tendency. If the figures for Montana be omitted there is practically no correlation either positive or negative between rates of increase in population and rates of increase in teacher preparation. If there is a tendency it is toward more rapid improvement in teacher preparation in the active states having large population increases.

Conditions of Progress

The causes operating towards the improvement of the education of the teaching force can probably be determined accurately only by a more detailed analysis of circumstances in particular states. Clearly enough the first step in any such analysis is a careful compilation of the facts as to teacher preparation from year to year and a thorough interpretation of the figures once they are compiled. At present data on the educational preparation of teachers are collected in less than one-third of the states of the Union and in only a few of these states are the figures subjected to careful scrutiny that their true import may be discovered. Notable examples of careful collection and use of such figures are found in recent school reports from New Jersey, Massachusetts, and Montana.

It is a truism in education that as is the teacher so is the school. The development of modern educational statistics in the past decade has taught us the value of child accounting, and we have learned that conditions affecting the progress of children are most rapidly improved by the simple method of checking up the facts each year with regard to the ages, the grades, and the progress of the children and printing them in our annual reports. We need to learn the same lesson with regard to the teachers and to develop methods of teacher inventories which shall annually gather and publish the facts that will tell us where the teachers come from, how well they are prepared, how long they serve, and what they are paid.

Summary

1. From 1910 to 1920 there is an increase of 42 percent in the average index number of teacher preparation for ten states.

2. In the years 1919 and 1920 little progress was made in teacher preparation, but on the other hand there was no substantial loss in any state for which we have figures.

3. When computed as a coefficient of regression the annual increase in years of college or normal-school training per teacher for ten states is 0.041.

4. If present rates of progress are continued until 1950, New Hampshire will lead the states for which we have figures with average teacher preparation nearly equivalent to four years of college or normal-school training.

5. There are no obvious explanations for the very diverse rates of progress among states. The situation clearly calls for careful collection and analysis of statistics in each state.

SCHOOL VARIATION IN GENERAL INTELLIGENCE

WARREN W. COXE
Formerly Vocation Bureau, Cincinnati, Ohio

A course of study which shall be of universal application throughout a school system has certain obvious advantages, among which are the following: 1) inasmuch as all pupils of the same grade study the same material at the same time, transfers can be effected easily and without loss of time to the pupil; 2) there can be a greater uniformity of textbooks; 3) courses of study can be worked out in greater detail; 4) many supervisory problems are simplified; and 5) the cost of education is somewhat less. All these advantages are on the side of less work for the teacher and less cost for the community.

We should not forget, however, that student bodies differ greatly in general intelligence and therefore in their ability to pursue a given course of study. Were we to judge the efficiency of teaching by the degree of success with which pupils pass standard educational tests, we might be unfair unless we made allowance for such a factor as the general intelligence of the pupils. It is, therefore, very important that a school superintendent should know the composition of the student bodies in the several schools in terms of general intelligence.

This article presents data regarding the general intelligence of 24 sixth grades in 24 elementary schools in Cincinnati. The examination used was the Otis Group Intelligence Scale. It was given near the end of the year, was administered throughout by two people trained in giving such tests, and was scored by the same two people. The conditions of giving and scoring were thus uniform in all the schools. The results ought therefore to be strictly comparable.

The immediate reason for giving the tests was the selection of candidates for a six-year classical high school. The tentative basis of selection was a minimum intelligence quotient of 110.

In Table I is presented the distribution of scores for the 24 schools. The wide variation of scores in any one school is at once noticeable. Schools are not alike in this respect, for some manifest much more variation than do others. Reduction of this

TABLE I. DISTRIBUTION OF OTIS SCORES OF TWENTY-FOUR SIXTH GRADES

Scores	Schools																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
180–184	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	1	...	...	...
175–179	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
170–174	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
165–169	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
160–164	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
155–159	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
150–154	2	...	...	...	...	1	...	...	...	1	...	1	...	...	...	...	2	...	...	1	...	...	...	...
145–149	2	...	...	2	...	...	1	...	1	...	...	...	...	...	...	...	...	...	...	1	1	...	...	...
140–144	4	...	...	2	1	...	...	...	...	...	3	3	...	...	...	2	1	...	1	...	...	...	...	...
135–139	1	2	1	...	...	...	1	...	...	4	4	2	...	...	...	...	1	...	...	...	1	...	...	...
130–134	7	1	1	...	...	1	...	...	3	2	...	3	1	...	1	...	1	...	1	1	4	...	1	...
125–129	8	1	1	...	3	1	...	...	...	1	4	10	1	...	...	2	2	...	...	3	3	...	...	...
120–124	4	6	1	3	1	1	1	...	1	3	1	8	1	1	1	5	1	...	1	3	7	2	1	...
115–119	9	4	3	3	1	4	4	...	1	1	7	3	1	...	2	1	2	2	1	3	3	1	2	2
110–114	6	4	6	3	1	3	10	...	1	7	3	1	1	1	...	5	8	1	1	6	5	...	1	1
105–109	9	9	2	11	4	6	7	1	3	6	4	10	1	1	3	3	6	...	1	5	6	...	2	4
100–104	6	3	7	3	3	6	5	1	2	3	4	9	4	...	2	8	7	1	6	5	10	2	3	3
95–99	10	9	7	5	5	4	2	2	6	10	7	7	3	2	5	5	8	3	3	12	11	...	3	3
90–94	8	6	7	2	7	3	3	1	6	6	1	8	2	1	4	9	8	...	4	8	7	2	2	7
85–89	13	8	7	2	4	9	7	3	4	9	5	5	4	1	3	10	4	1	4	10	5	2	9	4
80–84	7	8	8	4	2	7	6	3	4	13	5	9	1	...	4	9	11	3	3	9	10	6	1	2
75–79	7	6	8	2	4	4	3	2	4	10	3	7	1	6	4	8	5	7	6	6	11	6	7	2
70–74	4	11	8	2	1	5	9	6	1	8	5	4	5	2	2	5	7	8	2	5	7	4	3	...
65–69	4	7	3	1	2	6	5	3	6	6	3	5	2	...	2	8	8	2	2	6	10	5	3	4

Scores	Schools																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
60–64	8	2	4	4	1	1	1	4	2	5	....	2	2	3	2	7	6	7	4	1	4	3	6	1
55–59	3	3	2	3	2	....	1	5	1	4	1	1	2	....	....	....	5	4	3	4	1	6	6	2
50–54	3	1	....	....	3	1	3	7	1	....	....	1	1	2	6	3	3	4	3	1	3	2	....	....
45–49	....	2	2	....	....	1	2	4	1	....	2	....	1	2	2	2	....	6	2	1	1	2	1	1
40–44	....	....	1	....	....	1	1	2	....	....	1	1	1	1	....	2	....	1	1	1	....	1	2	4
35–39	....	1	....	....	....	....	2	1	....	....	....	....	1	....	3	....	1	3	2	....	....	1	1	....
30–34	....	....	....	....	....	....	....	3	....	....	....	....	....	....	....	2	....	....	....	....	....	1	....	1
25–29	....	1	1	....	....	....	....	2	....	....	....	....	....	....	....	....	1	2	....	....	....	1	....	....
20–24	....	....	....	....	....	....	....	2	....	....	....	....	1	....	1	....	....	....	....	....	....	....	....	....
15–19	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	1	....	....
	126	95	80	52	45	65	75	52	48	99	63	100	37	23	47	96	98	55	51	92	112	48	54	41
Average..	99.9	89.5	87.8	98.0	91.5	91.3	88.7	61.8	90.3	91.0	98.9	100.1	82.6	76.4	79.3	86.6	88.7	67.5	81.3	91.9	92.5	72.1	80.1	82.7
Average Mental Age	13^{10}	13^{0}	12^{10}	13	13^{2}	13^{1}	12^{11}	10^{8}	13^{0}	13^{1}	13^{9}	13^{10}	12^{5}	11^{10}	12^{1}	12^{9}	12^{11}	11^{2}	12^{3}	13^{2}	13^{3}	10^{8}	12^{2}	12^{5}

variation through a reclassification of the pupils should make the work of teaching very much easier and more effective. In another study we found that the variation in mental age was much greater than the variation in chronological age. This means that there is a tendency to allow children to progress in school in accordance with chronological age rather than mental ability, and this partially explains the wide variation of scores just noticed.

But the fact we wish to emphasize especially is the extent to which schools differ in the type of their pupils. This is better shown by a comparison of intelligence quotients. Scores will vary with growth, but the intelligence quotient is presumed to be nearly constant over a considerable number of years. Because of this constancy, the intelligence quotient is used to classify children as dull, normal, bright, or superior.

In Table II is shown the percent of pupils above and below certain intelligence quotients. Compare, for example, Schools 1 and 18. In School 1, 75 percent of the pupils have intelligence quotients of 100 or above; 57 percent of 110 or above; and 25 percent of 130 or above. In School 18 only 22 percent have intelligence quotients of 100 or above; 9 percent of 110 or above; and none of 130 or above. School 12 has children somewhat brighter than School 1; School 8 has duller children than School 18. Between these extremes are all gradations.

An examination of the data of Tables I and II shows that the school averages do not conform to the normal probability curve. This is because the tables do not include as many schools from congested districts as from the better residential sections. Were all schools included, there would be a larger proportion with low averages and the variation would be increased. Furthermore, in none of the schools are the extremely dull included in this grade, for they have failed of promotion and are either to be found in the lower grades or in special schools.

A comparison of these facts with the type of community served by each school is instructive. Following are brief statements descriptive of each community:

School 1—High-class residences; a large proportion of successful business and professional men.

School 2—In a congested district; includes stores and some factories; some foreigners; almost entirely a laboring class.

School 3—Mainly residential, but near factories; laboring class forms a large proportion of the population.

TABLE II. PERCENT OF PUPILS ABOVE AND BELOW CERTAIN INTELLIGENCE QUOTIENTS

Intelligence Quotients	Schools 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
100 or above	75	54	44	71	60	55	60	19	62	56	71	77	54	39	40	54	54	22	55	60	54	25	39	39
Below 100	25	46	56	29	40	45	40	81	38	44	29	23	46	61	60	46	46	78	45	40	46	75	61	61
110 or above	57	33	21	56	38	28	43	6	29	36	56	61	35	22	32	38	41	9	24	34	33	10	22	24
Below 90	11	32	34	19	18	22	23	62	25	20	13	15	30	39	43	31	27	64	33	25	27	50	39	34
130 or above	22	5	3	13	11	6	7	0	15	12	21	24	8	0	6	5	6	0	8	7	11	0	4	0
Below 70	0	3	5	0	2	3	9	23	2	1	6	1	5	9	13	5	3	15	8	3	2	8	7	12

School 4—High-class residences; homes of college professors and professional men; a small part is very wealthy and another part is decidedly tenement.
School 5—An outlying district; high-class residences; homes of business and professional men.
School 6—Good residences; homes of business and laboring men; an orphan asylum is within the district and sends its children here.
School 7—Good residences; a great number of excellent apartment houses; homes of business and laboring men.
School 8—Colored children only; a very high-class colored neighborhood; entirely residential; a great many apartment houses.
School 9—Good residences; many home owners; largely business men.
School 10—Thickly settled residential neighborhood; business and laboring men.
School 11—Good residences; apartment houses; business and office men.
School 12—Rapidly growing community; successful business and professional men; high-class residences; very homogeneous.
School 13—Outlying district; almost in an agricultural region; a very stable population.
School 14—Good residential; many negroes; laboring class.
School 15—Semi-residential; near the railroads and river; largely laboring class.
School 16—Part of district is high-class residential, part is tenement; an orphan asylum is in the district.
School 17—Rapidly growing district; mainly residential; homes of factory workers.
School 18—Almost in the heart of the city; very congested; low-class tenements; some foreigners; run-down neighborhood; laboring class.
School 19—Outlying district; has a great many children from an agricultural neighborhood.
School 20—Outlying region; high-class residential; business and professional men.
School 21—Outlying region; very good residences; business and working class.
School 22—Semi-residential; part of district is tenement; foreigners; laboring class.
School 23—Outlying district; fair residential somewhat rural.
School 24—Congested tenement district; run-down neighborhood; partly industrial.

It is manifest that these schools represent widely varying conditions. Children attending School 18 have a very limited home environment and nearly two-thirds of them are classified on the intelligence scale as dull. Children attending such schools as 1, 4, or 12 are of a radically different type both as to environment and general intelligence, being much above the average. These varying conditions indicate that the course of study is a problem for each school, to be solved according to local needs.

Enough experimental work has already been done to suggest how this may be carried out. Whipple and Terman have shown that bright children not only can make rapid progress but can at the same time take an enriched curriculum. It has also been found that dull children can almost never make normal progress through the grades. We thus have three groups, the dull, the normal and the bright, each requiring some special adaptation of the curriculum.

We have shown that these schools varied in the general intelligence of their pupils and that they served communities of very different types. It is also to be noted that there is a correlation between the intelligence level of the pupils and the character of the community. Those with lower intelligence levels generally live in the more congested districts; those with the higher intelligence levels generally live in the outlying, residential sections.

This suggests how the school may adapt itself to its community. Consider, first, how such a school as No. 18 will plan its curriculum. It is evident that it will have to permit its children to progress more slowly than usual, possibly allowing three years to do two grades of work. It should provide some of those things ordinarily found in the home but lacking in most of these homes, such as games, parties, plenty of simple reading, chance to "tinker," etc. Hand work should be provided, partly for cultural reasons, but mainly as a preliminary to vocational training. Since these children can neither make the rapid progress of bright children nor progress as far in school, one would only cause them to fail by urging them to meet the more difficult requirements. Failure tends to breed discouragement and discontent. If, on the other hand, the curriculum is within their capacity and definitely helps them to be self-supporting and to fill a place in society, their self-respect will be maintained and they can become valuable citizens.

Schools having bright children can expect the home to care for much of the training. These children can do more than average work, both in quantity and quality. They can possibly do three grades of work in two years and at the same time have their curriculum enriched. We do not know as yet just how the work should be enriched. A large proportion of such children will enter the professions or fill responsible positions in the business world. These vocations need a very broad foundation and an

extended amount of training. Anything the school can do to give better preparation in less time will be of great service to the individual as well as to society.

It is not likely that any school will be made up of but one of our three groups. It will generally have all three of them. An analysis of the kind here attempted will show which groups are present and what kind of adjustment is necessary to make the school of most service to its community.

SOME FURTHER STUDIES OF GIFTED CHILDREN

Elizabeth Cleveland
Supervisor of Girls' Activities, Detroit Public Schools, Detroit, Michigan

Having established enough seventh- and eighth-grade classes for gifted children to serve most of the city of Detroit, and having arranged a satisfactory system of testing those who enter and of following up those who leave, we feel we have made a worth while beginning in the special training of the most promising. But we are still looking with curiosity and awe at the impressive group we have brought together, trying to discover "how they got that way" and what we had better do about it.

In the months between March and June, 1920 we made some special studies of our three "special advanced" classes, comparing them with a control group of normal pupils in the same schools as to health, nationality, home conditions, types of reading and recreation, amount of travel, and vocational and educational plans.

The health studies were made with 140 pupils and with an equal number of control pupils of the same age in the same schools. Although we were unable to arrange for a thorough physical examination of each pupil, we could at least, with the assistance of our physical training department, measure weight and height and make a rough estimate of general physical condition. Of course, compared with other pupils of the same grade, pupils of the special group are likely to be smaller, because they are younger. Many of them come from homes where health is intelligently looked after, and they show the effect of refreshing sleep, proper food, fresh air, and exercise. In comparing a group of gifted children with a normal group we had always received an impression of more complete physical fitness in the former. This is expressed in posture, in nervous control, and in a general look of contentment and well-being. There are fewer wandering eyes, fewer open mouths, fewer restless hands. Even taking into account the interest in work specially and skillfully adapted to their ability, these children show a poise and alertness that seem to be due, at least partly, to physical causes. According to the standards furnished by the Bureau of Education at Washington,

we found 114 of the special group against 82 of the control group to be within ten pounds of the proper weight for their height. The teachers' estimates of general health showed no striking differences between the groups. This evidence, so far as it goes, seems to refute the old idea that the brilliant mind is usually found in the unhealthy body.

In questioning the pupils as to nationality we went back to their grandparents. For 114 children in the control group there were 176 reports as to the nationality of grandparents and in the special advanced group 222.

The distribution of nationalities is shown in Table I.

TABLE I. NUMBER OF GRANDPARENTS OF INDICATED NATIONALITIES

Nationalities	Control Group	Gifted Group
American	97	90
English	13	28
Canadian	27	15
German	12	36
Scotch	17	23
French	3	11
Irish	5	16
Swiss	1	..
Polish	1	..
Swedish	..	1
Welsh	..	1
Newfoundland	..	1
Total	176	222

This report is, of course, not a safe basis of generalizing, yet it suggests some interesting lines of investigation. The English, Scotch, and Irish usually make a good showing and some teachers have attributed this to the common language. But the French and German make a still better showing and the Canadians not so good. In a general way we have always observed that the children of some nations are bright and some dull; that Russian-Jews are quick to learn and that Poles are slow. But what determines the fiber of a group that has lived for centuries in the same conditions? Is it the inner structure or the outward circum-

stances? If we knew we might begin to modify in some respects our complacent uniformity.

In another connection when studying a group of girls retarded for various reasons (other than mental deficiency) we had found the most marked difference from the control group in their fathers' occupations, indicating social rather than mental or physical causes for their retardation. Furthermore our first groups of gifted children, selected according to the judgment of teachers and principals, had shown a decided majority of pupils whose fathers were in responsible business positions or in the professions —pupils who belonged to the prosperous, or at least to the comfortable classes. Perhaps their good English, easy manners, and general sophistication impressed their teachers as superior intelligence. At any rate we were surprised to find that in the present study this distinction tended to disappear. When the gifted children were selected according to the results of the tests the occupations of their parents were chiefly remarkable for their great diversity.

When it came to tastes and standards of living, as shown in the kind of reading and recreation and in plans for education there was a distinct difference between the groups. The pupils were asked to name their favorite books, which were classified as inferior, average, and superior. Dime novels, silly sentimental tales, the "Elsie" and "Polyanna" sort of books were considered inferior. Books read for information, books that appeal through subject rather than style, ordinary modern novels, were classed "average." Well-written history, poetry, the finest fiction, were called "superior." The results are shown in Table II.

TABLE II. DISTRIBUTION OF PUPILS ACCORDING TO THE READING PREFERENCES

	Control Group	Special Group
Inferior	49	10
Average	55	54
Superior	9	49
Total	113	113

It seems obvious enough that they appreciate because they are bright. But can it also be said that they are bright because they appreciate? Would they have passed the tests as they did had they not read so much and so discriminatingly? Certainly the brighter ones even at this age have the vision of higher education. Sixty-two of the gifted group were already planning for college. Perhaps also the intelligence is sharpened by experience of the world. Forty-eight of the gifted group against 36 in the control group had travelled considerably.

At the end of the year a report was made on the pupils who were found in any way unsatisfactory, with the teachers' comments on the cause. We found 22 out of 160 reported weak, 15 of these in one subject only, 14 of these 15 in Latin and 13 of them in one school. Fifteen of the 22 were in the VIIB (the beginning class), six in the VIIA, one in the VIIIB, and none in the VIIIA. Only four pupils were reported generally weak. The main cause of the difficulty in the case of 11 pupils was reported as lack of concentration and application, which of course need not indicate that the testing was at fault. The problem is to develop these powers.

There were only four whom the teachers considered lacking in ability. These will be referred back to the testing department for special study.

A follow-up report on 47 high-school pupils formerly members of gifted groups shows the following results:

Superior work (almost all 1's and 2's)	30
Satisfactory (2's or 2's and 3's)	13
Unsatisfactory (3's or 4's in more than one subject)	4

A report from a Latin class, six of whose members were formerly in gifted groups and most of whom had had the same teacher since the beginning of their course, showed the following averages as results of their mid-semester examinations.

Average of class	66
Average of five from gifted group[a]	80
Average of class exclusive of five from gifted group	62

All these reports are suggestive so far as they go, but much more observation and experiment is needed. We plan to continue the recording along the lines indicated above and particularly to extend our work in physical measurement. Our first

[a]One best student absent.

classes are just graduating from high school and we hope to follow the progress of each individual in college or in vocations. So that while we have made no progress in the remoter problems of determining causes, we can at least begin to measure the effects of our methods and make modifications accordingly.

MOTIVATED DRILL WORK IN THIRD-GRADE ARITHMETIC AND SILENT READING[1]

J. H. Hoover
State Teacher's College, Cape Girardeau, Missouri

The Problem

The problem of this study concerns the value of certain games or devices, to be explained later, in providing motivated drill work in the fundamental processes of arithmetic and reading. The study is based upon a few fundamental conceptions as to the nature and function of drill, i.e., of that kind of repeated activity which has for its purpose the increasing of one's physical skill or dexterity, or the permanent fixing in memory of certain useful associations. Drill, therefore, is an activity which has for its purpose the reducing of certain mental or physical operations to an automatic basis.

If drill is to be made effective and economical, it must be freed from some of its monotonous and unattractive aspects. When children see the need of the process, it will be interesting and impelling. This suggests utilizing the play instinct as evinced in games and dramatization. When drill is thus conducted it becomes something more than repetition; it becomes repetition *with attention*. Moreover, it addresses itself not merely to the group but to the individual. It is our purpose to show how this play instinct was utilized for drill purposes in arithmetic and reading and to indicate the results that were secured.

Arithmetic Materials

In order to realize this purpose, suitable material had to be devised. In arithmetic the drill work assumed the form of a game in which the four fundamental processes were separately involved. After a number of plans had been tried and rejected, the one finally adopted involved the preparation of sets of cards two inches by one inch. On these cards numbers were written according to the arrangement in the case of dominoes. For addition, subtraction, and multiplication a set consisted of 28

[1] This investigation was carried on with the cooperation of F. J. Kelley, Dean of the School of Education, University of Kansas.

cards displaying the combinations of a set of dominoes which extends to double 9, except that all combinations containing 0 and 1 and all doubles were omitted. A diagram of this set follows. The letter "A" merely serves to identify a particular set.

A $\frac{2}{3}$	A $\frac{2}{4}$	A $\frac{2}{5}$	A $\frac{2}{6}$	A $\frac{2}{7}$	A $\frac{2}{8}$	A $\frac{2}{9}$	A $\frac{3}{4}$	A $\frac{3}{5}$	A $\frac{3}{6}$	A $\frac{3}{7}$	A $\frac{3}{8}$	A $\frac{3}{9}$	A $\frac{4}{5}$
A $\frac{4}{6}$	A $\frac{4}{7}$	A $\frac{4}{8}$	A $\frac{4}{9}$	A $\frac{5}{6}$	A $\frac{5}{7}$	A $\frac{5}{8}$	A $\frac{5}{9}$	A $\frac{6}{7}$	A $\frac{6}{8}$	A $\frac{6}{9}$	A $\frac{7}{8}$	A $\frac{7}{9}$	A $\frac{8}{9}$

FIG. 1. CARDS USED IN ADDITION, SUBTRACTION AND MULTIPLICATION. CALLED SET A FOR REFERENCE CONVENIENCE

Since the cards shown in Figure 1 are not adapted to short division others were devised for that purpose. The aim was, that the numbers and combinations of numbers, should be exactly divisible by as many of the numbers below ten as possible. Accordingly, two different kinds of card sets were devised called Sets B and C. The number 35 and its first 9 multiples formed the basis of the first set; and the number 72 and its first 9 multiples formed the basis of the second. For drill in division by 5 or 7,

A $\frac{35}{70}$	A $\frac{35}{105}$	A $\frac{35}{140}$	A $\frac{35}{175}$	A $\frac{35}{210}$	A $\frac{35}{245}$	A $\frac{35}{280}$	A $\frac{35}{315}$	A $\frac{70}{105}$	A $\frac{70}{140}$	A $\frac{70}{175}$	A $\frac{70}{210}$
A $\frac{70}{245}$	A $\frac{70}{280}$	A $\frac{70}{315}$	A $\frac{105}{140}$	A $\frac{105}{175}$	A $\frac{105}{210}$	A $\frac{105}{245}$	A $\frac{105}{280}$	A $\frac{105}{315}$	A $\frac{140}{175}$	A $\frac{140}{210}$	A $\frac{140}{245}$
A $\frac{140}{280}$	A $\frac{140}{315}$	A $\frac{175}{210}$	A $\frac{175}{245}$	A $\frac{175}{280}$	A $\frac{175}{315}$	A $\frac{210}{245}$	A $\frac{210}{280}$	A $\frac{210}{315}$	A $\frac{245}{280}$	A $\frac{245}{315}$	A $\frac{280}{315}$

FIG. 2. CARDS USED IN SHORT DIVISION WHEN DIVISOR IS 5 OR 7. CALLED SET B FOR REFERENCE CONVENIENCE

the set based upon the number 35 is used and for drill in division by 2, 3, 4, 8, or 9 the set based upon the number 72 is used. If we attempt to use but one set of cards basing it upon 2,520, the

A 72/144	A 72/216	A 72/288	A 72/360	A 72/432	A 72/504	A 72/576	A 72/648	A 144/216	A 144/288	A 144/360	A 144/432
A 144/504	A 144/576	A 144/648	A 216/288	A 216/360	A 216/432	A 216/504	A 216/576	A 216/648	A 288/360	A 288/432	A 288/504
A 288/576	A 288/648	A 360/504	A 360/576	A 360/648	A 360/432	A 432/504	A 432/576	A 432/648	A 504/576	A 504/648	A 576/648

FIG. 3. CARDS USED IN SHORT DIVISION WHEN DIVISOR IS 2, 3, 6, 8, OR 9. CALLED SET C FOR REFERENCE CONVENIENCE

least common denominator of 2, 3, 4, 5, 6, 7, 8, and 9, the numbers are entirely too large for third-grade children to handle. Figures 2 and 3 show the cards as adopted. It would be better if the numbers were smaller and but one set were used; but to bring about this degree of simplification seems to be impossible.

Reading Materials

As in arithmetic, the aim in devising reading materials was to have drill work in reading assume the form of a game. In this, the elements of comprehension and speed were to play a prominent part.

Since love of activity is one of the characteristics of childhood, this fact was used in developing the reading materials. Printed cards containing "action" sentences which lend themselves readily to dramatization in the school room were devised. In determining the content of the sentences, the environment, interests, and every-day activities of children as a whole were kept constantly in mind. In order to appeal to the needs and interests of various types of children, four different kinds or sets of cards were devised which, for convenience, we will call, Sets A, B, C, and D. For a

description of these different sets of cards see, "Rules for the Reading Game," page 206.

A 125 School closes at four o'clock in the afternoon. Show how the face of a clock looks at that time.	B 67 A donkey was loaded with salt. He laid down in the water. What happened to the salt?
C 239 Shetland ponies are little horses which children like to ride. Show how tall a Shetland pony is.	D 87 Mosquitoes are larger than elephants. Their wings are made of brass and copper.

FIGURE 4. SAMPLE CARDS IN READING

A sample card from each of the four sets of reading cards is shown in Figure 4. Each card is 2 inches wide and 4 inches long. The A,B,C, or D, as the case may be, which appears on each card indicates the set to which the card belongs. There are 150 A's, 150 B's, 250 C's, and 100 D's. The cards in each set are arranged in order of difficulty (least difficult first, most difficult last), and the number of the card indicates its position in the set. For instance, A:125 means that the card belongs to Set A and is the one hundred and twenty-fifth card (based upon the author's judgment) in Set A from the standpoint of difficulty.

Set A is a group of "Action Cards." These cards are primarily simple commands or requests. The child works with things actually present, and no pretense is involved. Bodily activity is required in each case—e.g., place your right hand on your left knee.

Set B is a group of "Language Response Cards." Response to these cards is made wholly through the medium of spoken or written words. (Name some good winter games.) Bodily activity is not required. Language responses may be written or oral, depending upon the teacher's judgment as to the needs of the particular group of children in question.

Set C is a group of "Pretense Cards." Here the children are asked to *pretend* that they are doing this or that particular thing.

They work, or pretend to work, with things *not* actually present. (Act as if you were hoeing in the garden.) Muscular activity is required in all cases.

Set D is a group of "One Word Response Cards." Response to these cards may be made by using one of the four following words: *yes*, *no*, *right*, or *wrong*. For example, to the question *Is ten greater than nine?*, the response is *yes*. Again to the statement *Horses have two feet*, the response is *wrong*.

If a group of children need exercise in giving correct oral or or written language responses they should be given cards from Set B. If they need exercise in accurately getting the thought from the passage read so that they can perform the desired activity and thus give visible evidence of understanding or misunderstanding the passage read, they should be given cards from Set A or C. If they need practice in selecting the correct answer where other answers are possible, they should be given cards from Set D.

Rules for the Arithmetic Game

Addition.—Set A (numbers below 10) is used. The rules for playing this game are very similar to the rules for playing dominoes. Being adapted, however, to the needs of third-grade pupils. Children will learn to play the game more quickly if the teacher plays a game with one of the pupils and allows the others to watch while it is being played.

The pupils are arranged in pairs according to some convenient plan; for example, the child in the front seat plays with the one behind him, etc. Each pair of pupils is given a package of cards. For convenience let us say that Ruth and James are playing together. Each has pencil and paper on which to keep his own score.

James lays the cards on the desk, face downward, and shuffles them. They draw out seven cards apiece. Ruth holds her cards so that James cannot see them and vice versa. Ruth lays a card on the desk (say $\frac{4}{5}$) and adds the two *end* figures together. She puts "9" on her paper as her score for this play. From the seven cards in his hand James matches this card. Any card containing a 5 or a 4 will do (say $\frac{4}{7}$). Adding the two *end* figures (5 and 7) together he gets 12 as the sum, and puts "12" on his paper as his score for this play. Thus the game proceeds. If at any time Ruth cannot match the cards on the desk from the cards

in her hand she loses her chance to play and James plays two (or more) times in succession. Ruth plays again as soon as she can match the cards on the desk from the cards in her hand. There is no drawing from the unused cards. The game ends when one of the players no longer has any cards in his hand. They then add their scores, and the one having the *higher* total score wins the game.

The used cards are turned face downward and shuffled with the unused cards, then a new game begins. Playing is continued in this way until the time is called. Promptness on the part of the pupils in beginning and ending the play period will add much to the interest and usefulness of the game.

Subtraction.—Set A is used. The rules for "addition" apply to "subtraction" also, with the following exceptions; this time the end figures are *subtracted* (the smaller from the larger) instead of added. For the plays given in the above illustration, Ruth places "1" on her paper for her first score instead of "9" and James gets "2" instead of "12." At the end of the game the results are again added but this time the one getting the *lower* total score wins the game.

Multiplication.—Set A is used. The rules for "addition" also apply to "multiplication" except that this time the end figures are *multiplied* instead of added. Again referring to the illustration, Ruth places "20" on her paper for her first score and James gets "35." At the close of the game the scores are again added and the one getting the *higher* total score wins the game.

Division.—Set B (multiples of 35) or Set C (multiples of 72) is used. Sets B and C are very similar in purpose. They are designed for drilling pupils in the addition, subtraction, and multiplication of numbers *above* ten but they are *especially* designed for drilling pupils in short division and, to a limited extent, in long division. When the divisor is to be 5, 7, or 35, Set B is used. When the divisor is to be 2, 3, 4, 6, 8, 9, 12, 18, 24, 36, or 72, Set C is used.

The rules for playing differ in no fundamental way from those already given for addition. Children are arranged in pairs. Cards are shuffled, etc. For this particular day, say, drill in dividing by "4" is desired. The teacher writes the figure "4" on the blackboard where it can be seen by all. Again using our illustration; Ruth lays a card on the desk (say 72/288). Here the

end figures can either be *added* or *subtracted* depending upon the teachers' judgment as to which the class most needs—drill in addition or subtraction. In either case the result is *always* exactly divisible by four. If the *end figures* are *added* (using illustration above) the problem will be, 4 into 360, and Ruth will put "90" on her paper as her first score. If the *end figures* are *subtracted* the problem will be, 4 into 216, and the result is 54. The game proceeds until one of the players no longer has any cards. The scores are added, the one with the *higher* total score winning the game.

If drill is desired where the divisions are *not* always exact, use Set C when divisors are 5, 7, or 35; and Set B when divisors are 2, 3, 4, 6, 8, 9, 12, 18, 24, 36, or 72.

Rules for the Reading Game

The children are arranged in pairs according to some convenient plan. Each child is given a sufficient number of cards to occupy his time for the entire reading period. If the time allotted to a reading period is fifteen minutes, ten cards given to each child will probably be enough.

Suppose that Ruth and James are playing together. Each is given (say) ten cards from Set A. Each has a pencil and paper on which to keep the score of his or her opponent. James picks up one of his cards, reads it silently, and hands it to Ruth who also reads it. He then proceeds to perform the required activity. By his performance, Ruth judges whether or not he has the thought of the passage he has just read. She now gives him a score of "1," if he has performed his task correctly or of "0" if he has failed.

The teacher will do well to be in the midst of the children while the game is in process. She should watch the performances of the children who are being judged and the scoring of those who are doing the judging. Fairness, accuracy, and speed are to be encouraged.

Ruth now reads one of her cards and James becomes judge. Thus the game proceeds until the twenty cards are exhausted or until the reading period has ended. The one having the greatest number of perfect scores at the end of the play period wins the game.

The rules for Sets B, C, and D are the same as those for Set A. The only difference is in the nature of the response, and this does not affect the rules for playing.

Experimental Procedure

The experiment involving the use of these materials was carried out at Kansas City, Kansas. It included 1,139 third-grade children (571 in non-drill and 568 in drill sections) or thirty different third-grade rooms. The general plan of procedure was as follows:

1. At the beginning of the study (December 29, 1919) and at the end (April 1, 1920) standardized tests in arithmetic (Cleveland Survey Arithmetic Tests) and reading (Monroe's Standardized Silent Reading Tests) were given to all the children.

2. After the first tests had been given, an effort was made to divide the pupils of the thirty different rooms into two groups of equal size and mental attainments. In making this division the advice and assistance of the superintendent was sought. This method of division seemed advisable instead of waiting for the test results because it was desirable to begin the study without delay. The superintendent's judgment in this matter was exceedingly accurate; and this accurate division adds greatly to the value of the results of the study.

The division into groups accomplished; fifteen of the rooms were provided with the materials and the teachers were instructed in their use. The teachers of drill classes were requested to use the materials, both arithmetic and reading, 10 minutes a day on Mondays, Wednesdays, and Fridays. These instructions were carefully observed.

Since the aim was to compare the improvement of the drill group with that of the non-drill group, it was essential that the same amount of time be spent by each group in arithmetical or reading improvement. Therefore, the time spent in the extra drill work by the drill section was deducted from the regular amount of time given to arithmetical or reading improvement. In other words the time element in the two groups was identical, the only difference being the way in which this time was utilized.

In arithmetic the time allotted to the experiment was, as nearly as possible, equally divided between the four fundamental processes. No special instructions were given to the teachers of drill classes in arithmetic and reading (aside from the printed rules) except that they were to emphasize both speed and accuracy in the four operations of arithmetic and speed and compre-

hension in reading. They were to cover the same daily assignments in the textbook as the classes that did not use the drill materials. The teachers of the non-drill classes were asked to proceed with their classes in their usual manner, making no changes in their methods of instruction.

Methods of Scoring Papers

All the test papers were scored by the writer according to uniform methods. These methods followed the suggestions of the authors of the tests, except that in the case of the Cleveland Arithmetic Test, for which no accuracy score was provided, the writer used his own method of recording accuracy. This consisted simply in finding for each pupil the percent of correct answers on each sub-test, and in computing the mean of these percents.

The accuracy score for an entire class was obtained by adding together the accuracy score obtained by each class member, then dividing this sum by the number of pupils in the class.

For the methods of scoring speed in arithmetic, and rate and comprehension in reading, the reader is referred to the directions for scoring these items as found upon the score sheets which accompany these tests.

Results of the Experiment

The improvement of each section (drill and non-drill) as shown by the tests was calculated and put in tabular and graphical form (not shown in this article).

In every phase of arithmetic and reading considered in the study, the improvements of the drill section were more pronounced than the corresponding improvements in the non-drill section.

Let us first consider arithmetic. In Test A of the Cleveland Survey Test, the non-drill section changed during the study from a median of 10.9 to a median of 15.1 examples correctly solved in 30 seconds.

The corresponding improvement in the drill section was from a median of 9.7 to a median of 18.5 examples correctly solved in 30 seconds. Here the gain is 8.8 examples as compared with a gain of 4.2 examples in the non-drill section.

Similar comparisons of the gains (Tests A, B, C, D, E, F, G) made by the two sections (drill and non-drill) may be obtained from Table I.

TABLE I. COMPARISON OF ARITHMETIC GAINS MADE BY THE TWO SECTIONS

	Tests							Accuracy
	A	B	C	D	E	F	G	
Drill section......	8.8	5.7	6.1	7.0	1.9	1.7	1.2	17.8%
Non-drill section..	4.2	3.6	4.4	4.5	1.1	1.3	1.0	14.1%

In studying the subject of class gains two factors should be considered: (1) advancement along the scale of measurement of the median performance, and (2) the increase or decrease of class variability.

By referring to Table II the reader will note that in a large majority of the tests the *relative* variability in both sections (drill and non-drill) decreased during the study but that the decrease was more pronounced in the drill section than in the

TABLE II. COMPARISON OF ARITHMETIC VARIABILITIES IN THE TWO SECTIONS

A. NON-DRILL SECTION

Tests	January			April		
	Median	Quartile Deviation	Coefficient Variability	Median	Quartile Deviation	Coefficient Variability
A	10.9	2.6	0.24	15.1	3.1	0.27
B	5.2	2.3	0.44	8.8	3.1	0.36
C	2.9	1.0	0.35	7.3	2.3	0.33
D	3.0	1.1	0.37	7.5	2.5	0.33
E	2.6	0.8	0.31	3.7	1.1	0.30
F	0.7	0.4	0.59	2.0	1.2	0.60
G	1.1	0.5	0.45	2.1	0.9	0.43

B. DRILL SECTION

Tests	January			April		
	Median	Quartile Deviation	Coefficient Variability	Median	Quartile Deviation	Coefficient Variability
A	9.7	2.6	0.27	18.5	3.7	0.20
B	5.0	2.2	0.44	10.7	3.6	0.34
C	2.7	1.0	0.37	8.8	2.6	0.30
D	2.9	1.0	0.35	8.9	2.7	0.30
E	2.5	0.8	0.32	4.4	1.3	0.30
F	0.76	0.5	0.66	2.5	1.1	0.44
G	1.2	0.5	0.42	2.4	0.8	0.33

non-drill section. It will be noted also that the respective increases of *absolute* variability in the two sections are about equal. For example, in Test G the non-drill section changed from a quartile deviation of 0.5 and a variability coefficient of 0.45 at the beginning of the study to a quartile deviation of 0.9 and a variability coefficient of 0.43 at the close of the study. In the drill section the corresponding changes were from a quartile deviation of 0.5 and a variability coefficient of 0.42 to a quartile deviation of 0.8 and a variability coefficient of 0.33.

In reading greater improvements were also realized in the drill section. In comprehension the drill section gained 5 units as compared with an increase of 3.1 units in the non-drill section. In rate of reading the drill section gained 21.2 words per minute as compared with a gain of 12 words per minute in the non-drill section. Table III shows the details.

TABLE III. COMPARISON OF READING GAINS MADE BY THE TWO SECTIONS (MEDIANS)

	Comprehension		Rate	
	Non-Drill Section	Drill Section	Non-Drill Section	Drill Section
January.....	3.9	4.0	34.7	35.7
April.......	7.0	9.0	46.7	56.9
Gain........	3.1	5.0	12	21.2

In rate of reading the absolute variability (average deviation) increased for both sections, but the increase was larger for the drill section. The relative variability, however, owing to the larger gains in median performance was smaller at the end of the period than at the beginning. In comprehension of reading there was a slight decrease in absolute variability for the non-drill section. Both sections showed a decrease in relative variability. These facts are shown in Table IV.

TABLE IV. COMPARISONS OF VARIABILITIES IN READING

	Average Deviation			Coefficient of Variability		
	January	April	Increase	January	April	Decrease
Rate						
Non-drill section.	3.3	3.7	0.4	0.84	0.53	0.31
Drill section.....	3.1	4.5	1.4	0.78	0.50	0.28
Comprehension						
Non-drill section.	19.4	19.0	—0.4	0.56	0.41	0.15
Drill section.....	17.7	21.0	3.3	0.5	0.37	0.13

Suggestions Afforded by the Study

1. Uneconomical methods of drill are now being employed in the lower grades of our public schools.

2. Greater use should be made of the doctrine of interest, especially as it applies to drill work. Drill work should be motivated or vitalized by being connected with some dynamic purpose.

3. Bodily activity (dramatizations, handling of objects, etc.) can be profitably connected with achievement in school subjects.

4. Drill, to be efficient, must be made individual in character. It should be conducted, as nearly as possible, according to a child's needs and particular abilities.

5. Intensive focalization in connection with attentive repetition is an essential characteristic of efficient drill work, and by appealing to the play instincts of children this desired characteristic is effectively provided.

SOME ELEMENTARY STATISTICAL CONSIDERATIONS IN EDUCATIONAL MEASUREMENTS

J. Crosby Chapman
Yale University

The determination of norms of achievement for educational tests is admittedly a matter of great labor. It is the privilege of intelligence to analyze the exact advantages which are derived from any activity, especially when the activity degenerates into monotonous toil.

The question as to what are the exact uses to which norms of achievement are put can only be answered by considering the more general question. For what precise purposes are standard tests administered to a group? Two purposes are at once apparent: (1) to determine roughly the extent of the individual differences within the group; (2) to compare the achievements of the individuals of the group with those of other individuals external to the first group.

Obviously for the first use norms of achievement are superfluous. The differences within the group are the point of interest. Ninety-nine times out of every hundred, when a test is applied the major interest, if not the only interest, resides in the individual differences between the different members of the group. Even where an interest may extend to a comparison of the performance of this group with country-wide standards, such comparison is usually of little avail for reasons to be stated later in the article. In many cases we may be sure, therefore, that it is time wasted to secure norms based on large numbers. No tests which have been given sufficient trial to insure their validity should be held back because of lack of norms. The first point which demands consideration before attempting to collect standards on any test is: Is this test in its present form worth the time and energy that will be needed to collect a reasonably accurate norm? No physicist would dream of determining the constants of an instrument with great care, unless he was reasonably sure that the instrument was to be put to uses where such precise knowledge of constants would be required.

First, let us consider the question of the determination of norms and later we may take note of some considerations with reference

to construction of the scales themselves. It seems to have been accepted without thought by a number of educational workers that the values derived from the application of a test to several thousand children are of necessity more accurate and more valuable than the results obtained from a smaller group. There exists a pathetic trust in the saving power of large numbers. It is apparently thought that a norm is made much more accurate by increasing the number of cases from one thousand to seven thousand, and some authors have gone so far as to quote with pride that the norms are based on more than sixty-eight thousand cases!

It is fairly easy to see how the misconception has arisen that a mere increase in the number of cases must, of necessity, yield more accurate results. Our knowledge of the law of error has been evolved from a theory of probability, based on ideal conditions. We all have a nodding acquaintance with the perfectly homogeneous disc or coin, which mathematicians toss, and with the urns from which are drawn the gayly colored balls. Under such conditions as those found in the tossing of the coin, for example, it is obvious that as the number of cases is increased, the ratio between the number of heads and the number of tails tossed becomes increasingly nearer to unity, which we judge by experience is the ideal value to which the ratio is gradually approaching.

The reason why any considerable increase in the number of cases gives a more accurate value of this ratio is found in the fact that the successive tosses are performed under a definite and limited set of conditions. The first five tosses, we may say, form part of an infinite series, the series being homogeneous in the sense that each member expresses the result of a toss, performed under conditions which do not change. With such series, any considerable increase in the number of the series tends to give a more accurate value for the ratio which is being determined. There is no limit to the approach of this ratio to unity. The series can be increased until human patience or human effort is exhausted. Every addition to the series tends to increase the accuracy, though after a certain point is reached the increase in accuracy becomes almost negligible. Increasing the length of the series is theoretically valid when the series is homogeneous, and is practically beneficial when the resulting increase in accuracy adds materially to the

value of the norm as a practical instrument, with a full consideration of the conditions under which it is to be employed.

To pursue the second point with reference to increased accuracy as a result of increase of cases, if the series fulfills the requirements of homogeneity, then the simple formula holds, and the probable error of the arithmetic mean of a series of observations is inversely as the square root of their number. This fact shows how easy it is to overrate the effect of multiplying observations. How neglible must be the added degree of accuracy, given by the addition of fifty-nine thousand cases to an original nine thousand cases! If the material is homogeneous, the above formula holds and the gain in accuracy is small compared with the experimental error incident to the collection of such data. If, on the other hand, the material is heterogeneous, the above formula is no longer applicable, and the general procedure of compounding becomes meaningless. The accuracy of the educational norm is not primarily controlled by the total number of cases and the use of the probable error formula is to most readers very misleading; the value of the norm depends almost wholly on the care with which the selection of subjects has been made. It is with reference to this very question that the greatest carelessness has been shown by educational workers.

If we refuse to be blinded by the large number of cases on which the standards of achievement are based and examine the conditions under which this mass of material has been "rounded up," the confidence in the reliability of the norm declines. Taking any one of the older tests, there is no doubt that the times at which the test was applied to any particular grade have varied all the way from October to June. Grade VI is of course Grade VI, whether it is measured in October or June, but the results obtained in October cannot be compared with those of June. We might suppose that when we run into thousands of cases, the final norm would at least give us the average result of February. But two assumptions are made here. The first is that the curve of improvement is strictly linear between October and June, which we know is not the case; and the second is that the tests have been equally distributed over the entire school year. But tests are usually employed to measure the attainments as a result of certain instruction. This means that more of the tests are given in the later months of the school year than in the earlier months.

Or the problem can be looked at from another point of view. Suppose the absolutely true values of achievement for the whole of the United States (whatever that may mean) in a particular test on the last day of the school year are

Grade V	Grade VI
9	18

Then we may assume that the result of each month of instruction is to increase the efficiency by one unit. Let us suppose therefore that instead of the norms being determined on the last day, they have been spread over a period of the last month and a half, a generous assumption. Then the average attainment of Grade VI will be approximately 17¼ which differs ¾ from the true mean. Yet compilers of statistics are determining the value of these indefinite measures (in our case 17¼) by using thousands of cases! To go on piling up the number of cases in the hope of getting accuracy, when such a fallacy as the above exists, is futile. One might as well send to the Bureau of Standards for accurate electrical instruments to determine the resistance of a bar, and then allow the temperature to vary in an uncontrolled way anywhere from 20 to 100 degrees centigrade. Nothing further need be said with reference to the fallacy of selection and the actual discrepancies in administration; the first is a large factor, while the errors introduced by the second factor are sufficient to make compilers more guarded.

It may be urged, however, that all the members of a series upon which a grade-norm determination is based have at least this one factor in common, namely, belonging to Grade X. While this is of course true, the fact of belonging to Grade X is of little significance, seeing that it may mean, as we have shown, anything from member of Grade X in October to member of Grade X in the following June. Furthermore, Grade X has no magical constancy, even if we confine it to Grade X in June, or let us say the last week of the period spent in the grade. It is a notorious fact that grades vary enormously, even within the same school system. The variation becomes even greater when we include numerous school systems; and when we include the grades of the smaller school systems and rural schools, the factor of belonging to Grade X ceases to have any valuable significance. It seems a lamentable waste of time, when one essential element in the equation is this appallingly variable factor, to bother with one hundred thousand

cases, or even with the second thousand cases, in the hope of correcting the second or third decimal place in some other variable. As time advances, as educational procedure is altered, as the attitude to retardation and elimination fluctuates, the meaning of Grade X must always be unknown. The norms of educational measurement, whether by grade or by age, need only be determined in the very roughest manner, for determination by age is by no means free from blemishes!

It cannot be too clearly recognized that the combining of a large number of groups, each homogeneous within itself, into one large heterogeneous group, in order to determine a very significant norm, defeats its own purpose. The norm for each of the homogeneous groups may have been of interest and of value, but the combined norm is almost valueless. Thus, for example it is fairly significant to know that a particular test, tried out in a particular city on Grade VI children, during their last month in the grade, gave a median result of 18. But when these data are combined with the results of a very inferior school system and even of rural schools, the composite value becomes meaningless. Mongrel results of this kind are like mongrel animals, of little worth. Each of these values has significance, but when compounded they becomes noninterpretable.

The best plan therefore is to keep the distributions separate. It would be far better for any standard test to be accompanied by the actual distributions, determined at dates specifically stated, on groups specifically described, than to attempt to determine a median for one million cases made up of diverse groups. Even if we can assume that in future an attempt will be made to have the test standardized for the close of the school year, it will probably be beneficial to publish the norms, obtained from twenty or thirty representative schools, rather than to combine these norms into a single figure.

Another consideration which enters with the passage of educational measurements from its infancy is the factor that the results of a particular year cannot be compounded with those of the preceding year, unless nothing has happened during those years to alter the conditions of learning, etc. If new methods are being used, if the effect of giving tests has been to alter attitudes, to improve instruction, to increase the emphasis here, to eliminate the useless there, then obviously to combine the results is to

break the law which admits of such compounding. As Venn says, "There is a familiar, practical form of the same error. . . . It is that of continuing to accumulate our statistical data to an almost indefinite extent of time or space. If the type were absolutely fixed, we could not possibly have too many statistics. The longer we chose to take the trouble of collecting them, the more accurate our results would be. But if the type is changing, if, in other words, some of the principal causes which aid in their production have in regard to their present degree of intensity strict limits of time or space, we shall do harm rather than good if we overstep these limits." Thus, for example, to compound the norms which Ayres publishes for his spelling test with present-day norms would be to commit the above fallacy.

It may be urged that there can be no objection if a person wishes to determine what is the average median performance of the whole country. Certainly not, especially when the individual in question pleads academic interest in the national norms for their own sake. But we may reasonably ask, apart from the satisfaction of this peculiar interest, to what conceivable use this norm, acquired with great labor, can be put?

Very accurate work in the collection of norms is futile until the following details are known with regard to the members of the group: (1) chronological age; (2) mental age; (3) sociological status; (4) hours devoted to subject. While some of these facts can be quantitatively measured, others equally essential can only be roughly estimated. A norm can never rise above its origin. Shacks are suitable if one builds upon sand, but lofty edifices demand deep foundations. We must always remember that we are measuring human traits, the conditioning factors of which are most complex; and we must estimate our errors and modify our procedure accordingly. Until we have much greater insight into what constitutes random sampling, let us beware of wasting our time.

We may now consider another phase of educational measurement, the neglect of which is threatening the effectiveness and speed of growth of the test movement. Obviously the school situation demands tests and yet more tests, tests that are sufficiently accurate, not to reveal the acumen of the men who construct them, but for the practical uses to which they will be put. This in itself should stimulate the output of tests which provide rough estimates of traits that need measurement. One is slow

to blame a scientific worker because he is too careful, but he may be correctly blamed if he breaks his own scientific code. It is perfectly legitimate for the scientist to say, "I refuse to meet this practical demand," but he must obey certain well-defined rules or he ceases to be a scientist. One of these elementary principles of scientific method often overlooked is the important one of balancing the errors within an experiment. In one part of the experiment we must not attempt to get accuracy to one-tenth of a percent, when in another part of our experiment there is an obvious error of 3 or 4 percent. Educational statisticians are most guilty of this time-consuming, fallacious procedure.

An illustration of a failure to balance errors which has gone unchallenged may be taken from Trabue's valuable study of completion tests. It will be remembered that the author wades through an enormous amount of statistical calculation in order to determine with slightly greater accuracy the relative difficulties of the various sentences on his scale. When has his scale been used for a purpose which demanded this refinement? However, ignoring this practical consideration and confining ourselves wholly to the scientific technic involved, it may be pointed out that the assumptions which have been made by the author are so great that any attempt at extreme accuracy is absurd. Note a few of the assumptions made: (1) that language completion ability is represented by the much abused probability curve; (2) that the variability of the different grades is equal; (3) that the time allotment was sufficient to give a "reasonable amount of time for each sentence." One almost hesitates to think what would be the effect on the elaborate superstructure of individual values of each sentence had they been given with different time limits and in a slightly different order. In spite of these assumptions, the writer estimates his intervals between grades by three methods, working in each case to the third decimal place and in some cases actually to the fourth decimal place! Incidentally I may remark that my own work shows the same weakness. We have all sinned. Most of us have strained at gnats and swallowed camels. Possibly swallowing camels is the fate of the pioneer, but it is a pity for the output of tests to be small because of the gnats.

Similarly the calculation of regression equations for purposes of weighting, as in some of Kelley's recent work is often much more a measure of the degree of random sampling attained than it is a measure of the relative effectiveness of the various measures involved. As a method, the study is of value, but the results are reeds, painted to look like iron. Several recent articles show that there is danger of the partial correlation method degenerating into a fashionable fad rather than serving a limited practical purpose.

The same is true in our straight laboratory work. I well remember my surprise on being told by an instructor in a psychological laboratory when engaged in the star tracing mirror experiment, that I must use a stop watch and not my own watch in order to read to fifths or at least half seconds. I tried to point out that there was an error of a very high order in the reckoning of what constituted the necessity for a retrace, but it was all to no avail. Accuracy was necessary for its own sake.

In the majority of cases where correlation is employed, the fact of interest is usually whether the correlation is high or low, yet how often the author, with apparent desire for intellectual exercise for its own sake, will determine the relation by both rank and product-moment methods, when one would be ample for the purpose in view. In many cases if judgment had been used it would have been seen that the correlation formula itself could hardly be applied under the circumstances. A friend of mine, when engaged on a piece of work which involved about two hundred correlations, had occasion to show it to some advisors. He had used the quick foot-rule method of correlation (his data made the use of any formula rather doubtful), and in any case he was interested merely in big differences. Yet with one exception his advisors suggested to him that he use the more refined formula of product moments. On pressing them for ultimate purpose they all admitted that their reason for suggesting it was based on habit and not on analysis. I am afraid the publication of correlation data to the second place when the total interest is in the first figure, and where the assumptions and shortcomings of the experiment have nullified the accuracy of the first figure, let alone the second, has had a bad effect on the scientific conscience. No corrections for attenuation can balance limited or slipshod selection of cases.

I have not considered the time consumed by many writers in perfecting a test in one direction while in other directions there are glaring practical deficiencies. Nor have I stressed the fact that within my whole experience I have never met a situation where anything but a rough instrument was needed. Perhaps one of the causes of this false striving for refinement may be found in the fact that Columbia University, which has so ably led the way in this type of measurement, has produced its studies for the double purpose of training immature scientists in methodology and of meeting a practical demand. The first factor, combined with the eagerness of educators to establish the dignity of their new-born science, has resulted in the latter being dressed in the garments of the more refined sciences. Too many clothes impede an infant science. Great care must be taken that the training in accuracy is not accuracy for accuracy's sake, but balanced accuracy which contributes to the attainment of a scientific goal. Those who find pleasure in very refined measurements for their own sake are prostituting their talents in the pioneer work of educational measurements. For keen analysis there is crying need in education but for ultra-refined measurement there is a more imperative demand in other fields; in these more settled regions this rare inclination will find ample scope and great reward.

Editorials

VARIABILITY

A bit of description in a recently published short story contains the following passage: "He was a tall man, with a red face, a large nose, fat cheeks, and a pendulous lip. His arms were long, etc., etc." Perhaps this is enough of a rather disagreeable picture; and, as in the story, so here, we shall skip to the next paragraph.

Such a passage really does evoke a picture, if we stop to let it. It has meaning. We are led to inquire how this comes about. In particular, how red must a face be to be called red, or how fat may one's cheeks be and yet escape the epithet? Under what circumstances are lips pendulous or arms long? How large is a large nose and will the same dimensions when applied to an eye or an ear justify the same adjective? Or, to proceed to extremes, would they justify it if applied to a man? This, you will say, is absurd. A nose three inches long is undeniably large, but a man three inches long would be too small to be a man. Even a human hand three inches long would not be large, but, on the contrary, pitifully small. These things, though equal, are yet unequal. Though they have the same measurements, they are somehow large or small in virtue of something independent of their size.

The fact seems to be that each judgment we make is with reference to particular standards. Although we measure teachers' salaries and the cost of office furniture in the same units we judge salaries and furniture costs each by their own standards. Cheeks are fat and noses are large because of something we have organized into our experience about the fatness and largeness of these features. What is the character of this experience? What data does it provide? By what fusion of past impressions are we enabled to pass judgment upon objects and processes, upon abilities and characteristics, upon men and events, shaping in what we call a reasonable way our thoughts and actions with reference to them?

We maintain that the data by which these concepts develop are twofold. In the first place, we recognize consciously or otherwise a type or standard for each class of things; and in the

second place, we have a notion of the extent to which each thing may depart from its type or standard without becoming unusual and the extent to which it may do so without changing its class.

Now, there is a very real sense in which statistical method abbreviates this slow process of forming concepts. For the type or standard it offers the average or measure of central tendency; for the extent of departure from type it offers the measure of variability. In using statistical method, therefore, we simulate the natural processes of the mind. We gather, for example, measures of the heights of Anglo-Saxon men—men whom we have never seen and whom other people measure for us, vastly extending by this means the range and number of our observations beyond those which we could possibly make personally. By rigorous methods we determine the average height of these men, and we probably find it to be about sixty-eight inches. Owing to the large number of cases at our disposal and the precision of the method employed, this result undoubtedly gives us a far more accurate notion of the "man of average height" than mere fortuitous experience could ever afford.

But our judgment as to whether a man who is above or below average is tall or short, and as to whether he is moderately or excessively so, depends not only on how much taller or shorter he is than the average but also on the closeness with which the heights of all men are found to group around the average. What serves to distinguish an individual is not merely his deviation from the average, but in reality that deviation in relation to the *typical deviation* of men in general. In other words to know that a man is seventy-one inches tall and hence three inches above average does not permit us to call him a tall man unless we know that three inches above average is unusual. Now, as to what we have just called the "typical deviation," we should find on examining our data that half the men whose measurements we had secured were within about two inches of average height. Our line of thought, therefore, might be somewhat as follows: this man is three inches taller than the average; of men in general who are above average height, half of them exceed the average by two inches or less; therefore, this man is tall.

"Two inches" in this case is a measure of variability. In so-called normal distributions it is the median deviation—that deviation from the average which is exceeded in half the cases.

Applied to men of shorter-than-average stature (average being sixty-eight inches) it would suggest that those of less than sixty-six inches might be called short.

Of course, we may be so constituted that we withhold the adjective tall or short—or hot or cold, rich or poor, skilful or unskilful—unless the object possesses appreciably more or less than this average-plus-or-minus-the-median-deviation of the quality in question. This is a personal matter; and because it is so, we encounter that vagueness of meaning that so often interferes with the understanding of language. Nevertheless, it seems not unreasonable that, for example, a man over seventy inches (five feet ten inches) in height should be called tall while one less than sixty-six inches (five feet six inches) should be called short.

In the case of intelligence quotients, the average is 100. But much larger numerical deviations from the average occur than in the case of heights of men. The median deviation, instead of being something like *two*, is about *twelve*. This means that excess above (or defect below) the average must be numerically six times as great in order to be equally significant. In other words, *twelve* above average has the same meaning in respect to brightness that *two* above average has in respect to height; that *twenty-four* above average in the one case means the same as *four* above average in the other; that *thirty-six* corresponds to *six;* etc.

Since in this sense *twelve* units of brightness have the same meaning as *two* units of height, these amounts are given a common name (median deviation); and this median deviation is used as a common unit.

We may put this in schematic form as follows:

Average height of men = 68 inches
Median deviation (M.D.) = 2 inches
Average brightness = 100 I. Q.
Median deviation (M.D.) = 12 I. Q.

Assuming the accuracy of these figures, we may make certain statements, such as that a tallness of 72 inches (2 M. D. above average) equals a brightness of 124 points (also 2 M. D. above average), or that a tallness of 74 inches is three times a brightness of 112.

At first sight this sort of comparison may seem forced. Yet as a matter of fact it is very common. We lately heard it said of a man who smote an anvil on week-days, but who smote a pulpit

on Sundays to better effect, that he was a better preacher than he was a blacksmith. A man may be a better prose-writer than a poet or a better thinker than a doer. A teacher's training may be better than her experience, or she may be a better teacher of Latin than of sewing.

Indeed, it is continually necessary in daily life to compare things which, so far as they are measurable at present or conceivably measurable at some future time, must be expressed originally in units of different kinds—such units as dollars, pounds, years, words spelled, problems solved, and errors committed. Only by converting the number of these original units possessed by a person or thing into some unit of variability such as the median deviation (or the standard deviation or the average deviation) do we strike common ground. This assumes that we have collected many measurements like the particular ones which we are comparing and that from a distribution of them we have derived, either statistically or by the less formal process of conceptualization, a type or average on the one hand and a measure of variation from the type on the other hand. Thus a given child is provably a better reader than speller, because knowing what average reading and spelling are and the variability of children about these averages, we find that he stands better with respect to the reading average and in terms of the reading variability, than he does in respect to the spelling average and in terms of the spelling variability.

When we reflect as to what is meant by saying that a man is a better preacher than he is a blacksmith, we realize that even here we have a rough though by no means easily communicable notion of an average preacher and an average blacksmith, and that *as preachers run* (i.e., according to their variability) the man in question is judged to exceed the preacher average by more than *as blacksmiths run* he exceeds the blacksmith average.

Thus, the two factors both in statistical and conceptual thinking are *type* and *variation*. Indeed in all our communication of thought which exceeds the Scriptural "Yea, yea" and "Nay, nay" these two facts are implicitly present.

It is clear, therefore, that in the formal processes of statistical method we cannot properly dissociate these two complimentary data. In particular, we cannot rely on averages alone. Many just criticisms of statistical statements have been incurred be-

cause nothing but averages have been reported and used. An average tells nothing at all about the extent to which items are grouped about it. Hence, in judging about a given magnitude it merely constitutes a point of departure.

For example, two teachers each receive a salary of $1,200. One is in a city in which the average salary is $1,000 and in which no salary falls below $800 nor exceeds $1,200. The other is in a city in which, although the average is likewise $1,000, the range is from $500 to $2,000. If we only know the average, all we can say is that each teacher is getting $200 more than the average. With only this amount of information the conditions look alike. In reality the teacher in the first city is receiving a high salary—the maximum—while the teacher in the second city is receiving relatively a much lower one.

The essential basis for the really significant statement of the amount of departure from the average is the natural, common unit of variability. Moreover, this variability is just as truly a norm or standard as the average which is so generally and so exclusively regarded as such. In fifth-grade arithmetic, for example, it is as valuable to know the typical variability as it is to know the typical central tendency. A teacher is as truly concerned to know how much, in point of intelligence, pupils differ from a standard variability as she is to know how much they differ from the average. Indeed, it is perhaps reasonable to believe that she is more interested and more affected in her work by the diversity of talent than she is by its general level.

Since, therefore, in language and in statistics a measure of variability is essential, since it forms the basis of our judgments and enters into all our thinking, since it is the natural unit in which all magnitudes may be expressed and in terms of which they may be compared, since it is as truly a norm or standard as the average itself, and in short since for all these reasons it is the indispensable supplement of the average in characterising both groups and individuals, we urge that in educational reporting it be more systematically presented, that in reaching decisions it be more generally regarded, and that in teaching and administrative procedures it to be given its rightful place.

We feel sure that if research workers more generally report measures of variability, and if they more fully interpret them and show their significance, these measures will enter more fully

into the consciousness of school people to refine their thinking and to make their judgments more discriminating. Here, as everywhere, the real test of the efficacy of research work is to be found not merely nor chiefly in the extent to which it satisfies technical requirements but in the extent to which it meets the practical needs of the hour.

B. R. B.

"SELLING" EDUCATION

Education is a great enterprise. Locally it may appear small. The neighborhood schoolhouse, even though it may not quite be of the "little red" variety, is frequently unprepossessing. Yet altogether apart from the emotional notions of the uplifter and of the demagogue when he is after the support of teachers, education as an institution is big in a material sense.

One-fifth of the inhabitants of the country are attending school; and practically everybody else has either attended or will attend. It is the only establishment save government itself which assays to minister to all the people. Nearly a million persons are engaged in conducting its affairs. Not infrequently one-third to one-half of the proceeds of taxation are devoted to its support. It is a business whose offices are everywhere—not only in every populous city, but in every village and hamlet. Its holdings, in mere acreage, constitute a principality. Its output is trained boys and girls—a product upon which no value is set, because it is priceless.

Big, however, as education is, its size is not appreciated by the people. It is so near at hand and so modest in its local setting, that the general public fails to realize that it is, in fact, near *everybody's* hand and that it has as many local settings as there are localities. The schoolhouse and the school taxes, the school teacher and the school children are so much a part of our lives that we do not realize that all these things in their present form are distinctly new and peculiarly American. We fail to understand that substantially every square mile from coast to coast and from Canada to Mexico is part of a school district.

The people are entirely too prone to accept education as a matter of course. We have it where we are now living and we shall have it if we move elsewhere. We have it today, and we shall have it tomorrow. Why take thought, then, for this institution which is always and everywhere present?

To be sure, some of us know that education is by no means a matter of course. To us it is a complex and difficult organization—interesting, ever new and vital. Its problems are of far-reaching importance and their name is legion.

But we must bring this conviction to the people. We cannot sustain education by our own efforts. The most fundamental necessity is public support. None of the solutions of our problems can be put into effect without money.

With this in mind, and with the determination to keep faith with our cause and with ourselves, we must exalt education as an institution in the minds of the people. In the lingo of business—we must "sell" education to the people.

Nor should this prove to be difficult. We have a great cause. In it the people, though they may appear apathetic, have an abiding faith. Even a moderate degree of effort may be expected to produce large results.

Contrast, if you will, the task of selling education with the task of selling insurance or automobiles or patent medicine or chewing gum. Note the skill of the advertiser in creating among a large number of people wants more or less unnecessary and hitherto entirely unsuspected. Observe the adroit manner in which the salesman displays his wares. The selling campaign is conducted with a high order of business ability. We cannot withhold our admiration, although we may feel that this ability is often devoted to unworthy ends.

Education, on the other hand, is already half sold. To sell it completely is perhaps the easiest of selling enterprises. A little more than a year ago, we attended a conference on the educational crisis called at Washington by the Commissioner of Education. This conference was attended by several newspaper men who were more or less interested in education. It was their unanimous opinion that to arouse the people to the importance of education was, as one of them said, "a lead-pipe cinch." This particular speaker went on to show how, unlike other enterprises, education could count on a fundamental interest from the beginning; how, therefore, interest would not have to be created; and how wants would not have to be manufactured. Nevertheless he showed that the material which was being furnished to the press on the subject of education was ordinarily so poorly prepared as to be of doubtful value.

Believing, as we do, that almost if not quite the most vital need for the immediate future is a true sense of the value and dignity of education as a democratic institution, we urge the necessity of selling education throughout the country. The method is clearly that of publicity. If the interest of the people is real, though latent, we should arouse it by legitimate publicity methods until it becomes a compelling force. There is nothing more important today than that this work be done, and that it be well done. There is no lack of material; and if this material is properly presented, there will be no lack of available channels of publicity. Every group of teachers should have its publicity committee—not a committee for securing better salaries, but a committee for placing before the people the purposes, the problems, and the achievements of the public schools. Not only elementary and secondary schools but every form of higher education should be included. Every newspaper within the city, county, or state, as the case may be, should be reached with simple, interesting, and continuous copy. Alliance should be sought with the churches, with Rotary Clubs and similar organizations, with Chambers of Commerce, with agricultural societies, with parent-teacher associations, and with labor organizations. In fact wherever people are brought together in such a way as to exert an influence upon the public or a portion of it, there the publicity campaign should seek to be effective.

Nor should this publicity movement be, as it has been in the past, a sporadic movement, hastily gotten up, directed toward a narrow objective, and abandoned when a decision concerning the objective has been reached. It should be a continuing policy. Its organization should be essentially that of taking the people into the confidence of those who are carrying forward the work of training children and young people.

In this enterprise it is clear that information, accurate, reliable, and significant, must be available. Moreover, it must be presented attractively. Workers in educational research are precisely the persons who have, or should have, this sort of information. At present, too much of it is being kept in our files or communicated to each other without becoming widely known. We realize, of course, that our researches must eventuate in some form of record or report. Let us realize that in addition to this we have a far greater task to perform; namely, that of supplying the data, as well as of participating in the organization, for a campaign, continuous, far-reaching, and disinterested in *selling education to the people.*

B. R. B.

Reviews and Abstracts

ARPS, GEORGE F. *Work with knowledge of results versus work without knowledge of results.* (Psychological Review Monographs, whole no. 125, Volume 28, no. 3.) Princeton, New Jersey: The Psychological Review Company, 1920. 41 pp.

The monograph reports a study with the ergograph—an instrument which lends itself well to the investigation of the progress and effects of fatigue. Roughly, such an instrument consists of an arm rest into which the subject's arm is firmly strapped; and the work consists in either raising a weight or pulling at a spring by repeated flexing of a finger, to which a cord is attached. The purpose of the apparatus is to exercise repeatedly a single very restricted set of muscles. The arm rest prevents pulling at the cord by moving the whole arm or otherwise than by means of the flexor muscles of the particular finger experimented with. The finger works against a known resistance (in most of the present experiment the load was four kilograms); and in other ways the situation is controlled so that it is possible to study with a considerable degree of exactness the progress of fatigue and factors conditioning efficiency, as these factors appear in the repeated functioning of a single relatively isolated part of the neuromuscular apparatus. In the experiment here reported the right forefinger was used; and it was flexed once each second, as timed by a metronome. After every ten flexions there was a rest period, varying from 0 to 10 seconds. The experiments were made every other day; and each day the work was continued until the finger was "exhausted" —that is, failed twice in succession to move the weight. The very important and very practical question was whether there was any difference between the amount of work done when the subject knew his previous records and could observe his record-of-the-day as he made it, and the amount of work done when the subject knew nothing of his record. Three subjects in all took part in the experiment; and one of these subjects continued in the work three years, thus becoming very thoroughly habituated and trained to the procedure.

Several very interesting results appeared. In the first place (the major conclusion) the amount of work done averaged distinctly greater with knowledge of results than without knowledge of results. It is pointed out that "Will power as conventionally regarded is inadequate to explain the efficiency differences"; the subjects throughout were instructed to make maximal effort. Rather, knowledge of results operated in some fundamental way to be "provocative of greater functional changes in the central nervous system." Work without knowledge was reported by the subjects as very deadening, incapable of being sustained, lacking in vitality.

In the second place, however, in a series without knowledge of results, run *after* a series with knowledge of results, it was found that imaginal elements tended to sustain the activity in somewhat the same way that actual knowledge of current results had done. Under these circumstances the subject imagined the record of each individ-

ual lift, comparing the strokes of the recording stylus in successive lifts, and so compensated, in a way, for actual observation of results.

In the third place, when work has progressed close to the point of exhaustion "a curious phenomenon of sudden recovery appears." In other words, when the movements have almost ceased there is a sudden recovery to efficiency almost as great as at the beginning. Indeed, this sort of recovery may take place a number of times.

These experiments (which may seem to the average school man, at first thought, to have little other than a theoretical importance for professional psychologists) really have a very interesting bearing on many current educational problems. The findings are surely one more excellent argument for systematic practice exercises in the various school subjects with individual records of progress accessible to the children or (perhaps) kept by them. Apparently work "with knowledge of results" not only gives better motivation but operates in some more fundamental way to increase efficiency. The sudden recoveries are also of no little practical interest. Evidently even in very simple forms of work there is a variety of compensatory mechanisms, the situation is highly complex, and the nervous system is protected in a number of ways from actual complete exhaustion of the apparatus involved in any one activity.

The study is based upon a mass of data which have been analyzed in most painstaking fashion. It is typical of the best type of laboratory work—a type of work of which there has been all too little recently, as a result of distraction of effort into the more spectacular activities of "testing." It is to be regretted that there is not more "transfer" of such laboratory findings into current educational thought. For the most part such psychological investigations as this seem to escape the notice of the educationalists.

S. L. Pressey
The Ohio State University

Wilson, G. M., and Hoke, K. J. *How to Measure.* New York: The Macmillan Co., 1920. 285 pp.

This is the latest book in the field of educational measurement which has come to the writer's attention. The volume contains twelve chapters as follows: The New Attitude Toward Measurement; The Measurement of Spelling; The Measurement of Handwriting; The Measurement of Arithmetic; The Measurement of Reading; The Measurement of English Composition; The Measurement of Drawing; The Measurement of Other Grade Subjects; The Measurement of High School Subjects; The Measurement of General Intelligence; Statistical Terms and Methods; The Teachers' Use of Scales and Standardized Tests.

In each of the chapters devoted to a single subject a number of the tests and scales appear complete with directions for use. Standards of achievement or grade norms are given and numerous distribution tables and graphs illustrate the treatment of data. Each chapter except the first closes with a bibliography on the field discussed in the chapter. The volume is written in an easy running style which can probably be read for the most part by the average teacher with little difficulty.

The authors make two statements in the preface by which it seems fair to judge the contents of the book. 1. "The present volume is dominated by two main ideas, first, that the work in measurement should be handled more and more by the individual classroom teacher; and second, that the chief purpose to be served by standard tests is the diagnosis of pupil ability and pupil difficulties." From the first clause of this

pronouncement one would expect the book to be devoted to the task of teaching the teacher the use of educational measurements. There is much evidence that the authors endeavored to do this but it is also clear either that they could not resist the temptation to include data which were at hand though not needed for this purpose or that they had other purposes in mind. For example much of the material in tables 4, 11, 14, 17, 24, 25, and 41 would not have been needed if the main ideas above stated had completely controlled. These are data of interest to supervisors rather than to classroom teachers.

Having declared that "the chief purpose to be served by standard tests is the diagnosis of pupil ability and pupil difficulties," one would expect the book not only to show the teacher how to locate the children having difficulty and the type of difficulty of each, but also to be full of specific helps for remedial work. Considerable attention is given to teaching the teacher how to locate the children, and some little effort is made to show her how to diagnose their difficulties; but little aid is given on the most serious problem—that of remedial treatment. In fact, the authors state specifically more than once that "it is not the province of the present work to discuss methods in any extended way, but merely to show the use of standard tests." (The quotation is from page 71, but the same idea is expressed on pages 21 and 43). Has not the day passed when experts in any field may consider their task done when they have pointed out ills or the means of discovering ills? It is true that the headings "Remedial Instruction and Remedial Measures" occur four or five times and that some suggestions are given under the title "Using the Results"; yet it can be clearly shown that the major emphasis is upon the form of the various tests and the standards derived.

2. The second prefatory statement by which one should examine the book is: "The purpose of this volume is not a critical evaluation of all the available tests on different subjects, but a treatment of those tests which on account of their use, purpose and adaptability have been found to be most serviceable to the classroom teacher." Those thoroughly familiar with the field of tests and measurements will doubtless wonder at the omission of Greene's Organization, Minnick's Geometry, and others which have been on the market for some time. If these are ruled out under the criteria above, one may ask why Rice's Spelling Test is given in full and why considerable space is devoted to various other tests of doubtful value.

E. J. Ashbaugh

Ohio State University

Proctor, William M. *Psychological tests and the guidance of high school pupils.* (Journal of Educational Research Monographs, No. 1.) Bloomington, Illinois: Public School Publishing Company, 1921. 70 pp.

This monograph is based upon an extended study of the relation between intelligence scores of high-school pupils and their educational success, as indicated by school marks, teachers' ratings, elimination, etc.

The study is of broad scope and deals with the important educational problems involved, rather than with the technic of intelligence testing. While the method used and the conclusions drawn are admittedly tentative, the monograph contains a large amount of material which will prove extremely helpful to high-school teachers and principals. The investigation extended over several years, thus giving sufficient time to check up the test results by records of the later success of the pupils. The

work was begun in 1916-1917 when the Stanford-Binet scale was given to 137 pupils in the Palo Alto High School. These pupils have now been followed for five years.

In 1917–1918 and in 1918–1919, 955 pupils in five high-schools were given the Army mental tests. The results of these have also been followed up and correlated with school success and elimination. Correlations of both the Stanford-Binet and the Army mental tests with school marks were in all cases high enough to show the tests to be of considerable value for educational guidance. Pupils who have low scores tend to predominate among those who drop out. Average scores of pupils enrolled in certain types of subjects differ greatly from those enrolled in other subjects. Score limits are indicated for the tests used, below which pupils are very likely to fail in Latin, algebra, etc. As would be expected, school marks in the semi-vocational subjects show low correlation with intelligence scores. The conclusion is obvious. There are plainly a great many children who, because of lack of ability to think in abstract terms, cannot succeed in mathematics, Latin, or English literature, but who can do fairly creditable work in household science, shop work, etc.

The reviewer believes the author's findings along the line indicated above to be extremely significant. Clearly it is going to be necessary to work out the chances of intelligence of various grades succeeding in the different kinds of school work. Until we have such information, educational guidance will be out of the question.

Another important problem is that of shaping the instruction in a given subject to the abilities of the children. Since the range of ability in any class is likely to be very wide, schools are compelled to consider the desirability of sectioning their classes on the basis of intelligence. This practice is becoming common in the grades, and is no less important in the high-school.

The author's chapters on the use of psychological tests in educational guidance and vocational guidance are based upon considerable experience in carrying on this kind of work both with high-school pupils and with Federal Board students in the university. The data presented show in a very convincing way that the intelligence score is always worth considering in advising any pupil regarding his future educational or vocational work.

Extensive data are presented on the university success of the high-school pupils studied. It was found, for example, that school marks of first-year university students at Stanford correlated to the extent of 0.46 with Stanford-Binet tests of these pupils given five years previously. This is one of the longest range comparisons known to the reviewer, and shows conclusively that intelligence test scores have considerable permanent value.

The monograph is a careful study of a problem which is certain to occupy the attention both of psychologists and of school men to a far greater extent than has been the case during the past. The importance of the problem is being rapidly emphasized by the passage of laws in many states designed to retain all children under the jurisdiction of the public school until the age of eighteen. The public schools of the future must deal with the entire child problem. They can no longer set arbitrary standards and eliminate the pupils who are unable to work up to these standards. Intelligence tests are necessary to help the school in making this adjustment.

LEWIS M. TERMAN

Stanford University, California

News Items and Communications

This department will contain news items regarding research workers and their activities. It will also serve as a clearing house for more formal communications on similar topics, preferably of not more than five hundred words. These communications will be printed over the signatures of the authors.

Adapting American Tests for Use in China

William J. Lacy, one of our correspondents in China, writes interestingly on the adaptation which is being made in China of the tests we are using here in America. Mr. Lacy is the executive secretary of the Conference Board of Education, Methodist Episcopal Church, Yenping, Fukien, China. We reproduce his letter in full:

Your letter of April 8th has been waiting my return from North China where I have been spending two months in famine relief work. I thank you very much for your invitation for me to contribute some brief report of the educational research work in which I have been engaged for the JOURNAL OF EDUCATIONAL RESEARCH, and shall hope to send you something within a few months.

At present I am at work on an adaptation of Daniel Starch's Arithmetic Reasoning Test which has been put into Chinese and is being used in our higher primary schools here in this province. The higher primary school is the top half of our American Grade School so that the tests are being used in our grades corresponding to the American 5th to 8th grades.

My initial investigation was carried on a year ago this spring at which time I was able to establish score values for the various problems which I had put into Chinese in a revised form. This year I have had the tests given in about thirty different schools and when all the returns are in and I have tabulated the results I hope to have a fairly well established standard score for each grade. It has been necessary, of course, in adapting these tests to make certain changes in subject matter as our American money values have to be rather radically changed because of the complicated monetary system in use here in China. Just to show you some of the difficulties in making an arithmetic test where money problems are used—and we certainly want to use the money problems because money plays the greatest part in the thought of the Chinese—I will give you the different steps in the changing of a dollar.

Money is spoken of in this part of the country as big money and small money. Big money consists of the dollar silver piece as standard. A dollar silver piece will change into eleven dimes and from one to four pennies, owing to the prevailing rate of exchange on a particular day. Then each dime changes into from eleven to twelve—usually twelve—pennies, so that a big dollar is worth anywhere from 130 to 135 or sometimes 140 pennies. Now, when you go to put any of our American arithmetic problems which deal with money into the hands of your Chinese, you have a problem that is not easy to solve. We have therefore more or less used the small money standard which consists of ten dimes to the dollar, but your dimes are again changed into more than ten pennies, so that in any such problems you either have to deal entirely with dimes and half dimes or else state the rate of exchange for pennies that is to be used in solving the problems. This, of course, puts another element into the solution of the problems and vitiates the results when it comes to any comparison of the same problems used in America.

I found in my initial investigation last year that the median score made by the Chinese pupils in the adapted tests which I had made was a little bit higher for the

same grade than that for the Starch Test at home. On the other hand in any test where the time element is used the Chinese students are much slower than American students.

What I have written is merely to give you a brief suggestion of what I am working on at present. I hope some time this summer or in the early fall to be able to complete the compiling of my results and get out a full report of the work that I have done with these tests. I shall then send you a copy of this report, and you can use such parts of it as you choose.

Use of Intelligence Tests in the Mission Schools of China

Mr. Fred P. Beach of the Department of Education of the Fukien Christian University of Foochow, China, gives us the most complete statement we have yet received as to the function which intelligence testing has or may have in the peculiar educational situation confronting educators in China. From what he says it is entirely clear that this function is by no means less than it is in America. Mr. Beach, like Mr. Lacy whom we quote above, is an example of the highly trained expert whom the missionary service is now enlisting. These men are evidently engaged in a momentous undertaking—an undertaking to which in their enthusiasm and ability nobly corresponds.

We are sure that our readers will find Mr. Beach's letter of unusual interest. It follows:

The real situation as far as tests are concerned is about as follows: They will be, when really ready and well-tried-out, an invaluable aid to us not only in our school work but in all of our missionary effort. Some eight years ago Dr. Paul Munroe came through here and I had a chance to visit with him a little. One of his remarks has stuck in the minds of some of us, namely: that the missions cannot hope to succeed in their conversion of a nation by the use of unintelligent material. Or as I paraphrased it "We cannot hope to make St. Pauls out of defectives, delinquents and dependents." Now it is the truth that the early mission had to take anything that came to it; and amongst the lot was a great deal of the type of the dying-out-family. There is nothing so strong in China as the Family. There is little individuality as we know it, and the successful man in China must support his relatives. The richer he becomes the more relatives he takes on. You can see that a Chinese without a family is in a pretty bad social and economic way. In their straits they have turned, many of them, to the church. Now the problem put up to us has been the education of their children. To a large extent the missions have been compelled to adopt the tactics of running very expensive schools and of subsidizing through these schools, those of its own Christian children who were bright enough. This is not the ideal way—not very democratic, perhaps—but as long as the mission schools continue to teach the coveted "English" better than the government schools so long we will continue on this tack. Now you can see that on this basis of operation a great many dollars could be saved to the missions if it were possible to give intelligence tests beforehand to these students and not have to wait for a year's school work to show up the poor students—and even then perhaps have to yield to importunity and give the child a second year. On the other hand, it will be as useful to us as to you to be able to pick out the extra bright and give them a better course.

There are probably no harder working students in the world than our Chinese. Nevertheless, we are facing in our schools the same problem of "not allowing your studies to interfere with your college course," and it is a powerful stimulus to be able to tell the students in chapel that intelligence tests do *not* test moral character or laziness, and that it is perfectly possible for a man to have high intelligence (I. Q.) and yet fail at the examinations, as a man "cannot remember what he has not seen."

Now as to technique. To tell the truth we are mere amateurs in the beginning stages of experimentation. Knowing Terman's book the best, we have based our work mostly on his. It is obviously impossible for us to pay fancy prices for printed tests by the hundred or thousand, and one wonders if they are any the better for being put

up that way. We shall have to develop some cheaper way to give group tests. Of course there is no copyright law and only our consciences will prevent us from pirating anything we need. But on the other hand *our main problem* is not the finding of American material and translating it into Chinese. The Translation problem is as easy as can be. Any University or Normal in China will have at hand some returned student who can translate perfectly. The *real question is the validity for Chinese society of the American or European Tests.* This is our experimental task for years to come. Inasmuch as ones "system of knowledge" (Pillsbury) is only built up by experience in this world as it is, these tests obviously must test the ability as it has been used in its social environment, and the good test is evidently that which is new to the person at the age tested and yet which depends on a certain definite experience and growth. In other words, your relativity again.

It is not difficult to change the Standford Binet tests so that instead of "the engineer going faster with a heavier train" it becomes a ricksha coolie with more people in his ricksha. But I take it that our task is to test thousands of children all over China till we have coördinated our work and gotten a pretty reliable idea of what can be expected of Chinese youth and children in such a social environment as they grow up in.

Meantime we have more than enough opportunity, as the tests are desired by all our schools. Our university will give a course in the Psychology Department next semester in order chiefly to get a set of students at work on the tests among the school children. Meantime we shall watch you fight out the validity of the underlying principles. I believe the tests America is turning out are useful even now, not as an absolute guide to whether a student shall be dropped or not, but as an additional help confirming in a general way the poor work of the student. I think we can safely use them in addition to the entrance examinations for those entering the High School or College.

Inasmuch as our Educational Association is striving now to supplant the once popular but now largely disliked "Uniform Examinations" and is trying to put "Standard Educational Measurements" in their place, it is obvious that any fair comparison of country schools as against city schools should be able to say what the median I. Q.'s of the schools compared are. The country schools try to do as much as possible for slow students whereas the city mission schools are rather relentless in dropping students who get a little behind.

In conclusion, I have no doubt that the tests may develop some national strengths and some weaknesses. Still I am confident that they will bear out the accepted theory that the Chinese are one of the great dominant races of the earth. The old examinations were in effect a test of intelligence; not a full rounded one, and certainly not a moral test at all. But for the most part it is the children of those who succeeded in the old tests that produce the greatest percentage of good students in our schools. Amongst our own church children there is great variation. But fortunately among them are a few of high ability, thanks probably to Mendelian proportions; and from these the church will get some leaders and it has already. Now, however, we should like to be able to know them when we meet them.

You may be interested in studying over the following vocabulary problem. In a country where ideographs are used and where verbal memorizing is still largely used, is a vocabulary test from the dictionary an intelligence test or is it a "measurement of classroom products?" This is a matter of immediate concern and is being worked out in several places in China. Notably Pekin and Nanking. We are doing some work here and will try some more but haven't gotten very far yet. To quote Mr. Hocking, "One doesn't need to apologize for not being perfect."

Before closing up his letter Mr. Beach adds this postscript:

Since writing the above I have given Terman's 100 word English vocabulary test to 30 freshmen in our University. Rather a wild chance perhaps, but I wanted to see what they would do with it and how they would compare with American children. I found that the best knew 51 words while they ranged down to 32, 65 being the number for "Average adult" and 75 for "Superior adult" in America. This, of course, raises the same question, "How much is this a test of classroom product and how much is it a test of intelligence?" It would be necessary to try this on all freshmen in China to

get any relative information. But it does indicate that there is a new angle of approach to the teaching of English in the Middle Schools here and that such a vocabulary test may well be used as a classroom standard for the students. I see that I must go on and make the same test on upperclassmen. The freshmen above mentioned were all in the neighborhood of 20 years of age.

On Burgess' "The Education of Teachers"

The following comment by Charles Carroll of the office of the Commissioner of Public Schools of Rhode Island together with a reply by W. Randolph Burgess came to hand during the summer. Mr. Carroll speaks first:

The JOURNAL OF EDUCATIONAL RESEARCH for March, 1921, printed "The Education of Teachers in Fourteen States," by W. Randolph Burgess. May I not be permitted to comment briefly upon the methods pursued and the conclusions drawn by Mr. Burgess, in the interest of greater accuracy in the use of educational statistics?

1. On page 161 the author quotes statistics of the educational qualifications of Massachusetts teachers, without indicating clearly that the statistics do not include all teachers employed in Massachusetts. Generally the Massachusetts statistics on the educational qualifications of teachers have been for full-time teachers only, and have omitted Boston. The omission of teachers employed for parts of years only, or otherwise irregularly, should have a tendency to produce averages more favorable than otherwise, in accord with a fair presumption that those well qualified probably will have more regular or full-time employment than those not so well prepared for teaching.

2. In the table on page 166 the writer has taken for Massachusetts the number of full-time teachers employed outside Boston plus an estimate for Boston based on the advance report of the Commissioner of Education for 1920. The table on page 161 indicates a significant improvement for 1920 over 1914 in Massachusetts; perhaps an estimate for 1918 based on figures for 1920 should be discounted by at least 2½ percent, or one sixth of the almost 15 percent increase between 1914 and 1920.

3. In the same table the estimated figures for Massachusetts full-time teachers are compared with the Rhode Island figures for all teachers, an objectional procedure for the reasons indicated in paragraph 1.

4. The author's interpretation of the diagram on page 167 appears to be incongruous if he is really attempting to find a fair index of preparation for teaching, because he fails to distinguish "mere schooling" from "professional training." Massachusetts is accorded first rank because of having 17 percent of college graduates amongst its teachers. Rhode Island's unsurpassed 70 percent of normal graduates is ignored or disregarded. In other words, the author appears to have accepted the hypothesis that preparation for teaching is to be determined by years of schooling. If the hypothesis were correct, a medical doctor with four years of college training plus four years of university instruction in a medical school, should be four times as good a teacher as a graduate of a two-year normal school. No one should question seriously the superior preparation of the doctor for the profession of medicine; but it is no more logical to hold that the doctor's education has fitted him for teaching than to assert that two years of study in a normal school is one-fourth of the preparation for practicing medicine.

5. In computing index numbers, as shown on page 168, the author has ignored the fact that in Rhode Island the normal-school course at Rhode Island College of Education has been 2½ years. Were credit given accordingly, the index numbers for Rhode Island would be 2.352, and Rhode Island, even on the count of all teachers as contrasted with only full-time teachers as in Massachusetts, and on accurate figures instead of estimates for Massachusetts, would lead Massachusetts and all other states.

6. In Rhode Island graduation from college is not recognized as adequate preparation for teaching. For many years the standard requirement for professional certification of college graduates has stipulated six semester courses in the science and art of education supplementing the requirements for the bachelor's degree. Because of this standard minimum requirement, and because of the relatively large number of

college graduates teaching in Rhode Island who have pursued additional graduate courses in education on state scholarships leading to masters' and doctors' degrees, it would be reasonable to rate the Rhode Island teacher who is a college graduate as having had at least an average of 4½ years of prepatation beyond graduation from high school. Were credit given for this also, Rhode Island's index number would be 2.424.

Mr. Burgess may not have been familiar with either the length of the course in Rhode Island College of Education or the certificate requirement for college graduates. In correspondence preceding publication of his article, however, his attention was directed to the fact that percents for Rhode Island based upon a full count of teachers probably would not insure full credit for the State.

Charles Carroll

Mr. Carroll's comment was sent to Mr. Burgess who submitted the following statement by way of rejoinder:

A complete answer to Mr. Carroll's letter would involve a detailed discussion of the records and computations back of my article in the March number of the JOURNAL OF EDUCATIONAL RESEARCH which would not be interesting to most of the readers of this publication, and would tend to be contentious. The reader who is interested in doing so may form his own conclusions by carefully reviewing my article in the light of Mr. Carroll's letter. The questions which he raises had all been carefully considered before the article was completed, and concerning most of them specific comment is made in the article. The state reports from which figures were taken are accessible.

I cannot help feeling, however, that Mr. Carroll's real difficulty does not arise from a difference of opinion as to specific details of my article. I think it rather arises from a difference of view point concerning the major purpose of the article. In the past few years statistics have become more practically useful than ever before. The reason lies largely in the wider development of the method of sampling, or representative returns. We are now able to measure price changes by a price index number which is based, not on a tabulation of all prices there are, but on a selected few. We are learning to measure human intelligence, not by tabulating all of the acts of an individual, but by studying sample behavior. In my article I wished to suggest a possible method by which we might secure a widely practicable measurement of teacher preparation. Completely exhaustive measurement of preparation is not possible. It is possible, however, to measure the situation by the method of sampling. This statistical method has always been open to the criticism of the person who calls attention to the specific cases which the method does not measure, but those who have experimented carefully with the method know that it, nevertheless, may give substantial truths. It is clear that some method of measuring teacher preparation is greatly needed. I think that the method suggested in my article would yield, with some further refinement in the collection of data, exceedingly valuable results. Even with the imperfect data now at hand the method yields results which are significant.

W. Randolph Burgess

On the Assumption That Errors of Estimate Are Equal in Narrow and Wide Ranges

Professor Kelley's statistical article in our May, 1921, number evidently evoked some interest—a mild and entirely decorous interest as would become those who are statistically minded. We received a comment on this article (or shall we call it a criticism?) from Doctor Karl J. Holsinger of the University of Chicago. He has located a really vulnerable spot in Professor Kelley's article, namely the assumption that errors of estimate are the same over wide and narrow ranges of talent.

This assumption, it will be recalled, occurred in the course of an argument to the effect that unless we know something about the spread of ability in the subject and relative series we cannot interpret the correlation coefficient. Professor Kelley's

specific statement here was that a correlation coefficient of +0.90 for children of the fourth to the eighth grades taken as one group might mean no closer relationship between the traits in question than a coefficient of +0.40 for children of the fourth grade only. Independently of the appropriateness of the specific criticism which Doctor Holzinger makes, this point is exceedingly well taken. It is worth repeating: we cannot know the real meaning of a correlation coefficient until we also know something of the range of the paired measures which enter into it. To this we judge that Doctor Holzinger, as well as every other thinker along statistical lines, will agree. Nevertheless, his position is well taken. He writes:

I have been much interested in an article by Professor Kelley on The Reliability of Test Scores, appearing in the May number of your JOURNAL. There are some points in the article that appear to be at least debatable—in particular the basic assumption that "the two errors of estimate are the same" on narrow and wide ranges (bottom of page 377). It would seem to be equally plausible to assume that the above errors are proportional to the corresponding "spread of talent" as given by the "true" sigma's. Under the second assumption, of course, r and R will be equal if the factor of proportionality is the same for errors and true sigmas. In other words, if we assume $\Sigma_{1.t}=k\sigma_{1.t}$, we have

$$\frac{k\sigma_1}{\Sigma_1}=\sqrt{\frac{1-R}{1-r}}$$

and for formula (2)

$$\frac{k\sigma_t}{\Sigma_t}=\sqrt{\frac{r(1-R)}{R(1-r)}}$$

Whence, if $\Sigma_t=k\sigma_t$, $r=R$.

Of course, all this is self-evident; but it shows the violence of Professor Kelley's hypothesis (cf. $r=.4$, $R=.914$, page 374). Spurious correlation due to age factor and heterogeneity will increase correlation for a large group such as Professor Kelley describes, but it is doubtful whether his formula properly accounts for such increase.

I am writing to ask you if you will give your readers the benefit of some editorial comment on some of these points, if you think it worth while. This matter of reliability is so important that a critical review such as Dr. Kelley's is most helpful, and might well be followed up by such comment as I suggest. I wish to express my high regard for Professor Kelley's work in general and in the present article. It is because of this that I am particularly anxious that we should agree on his methods.

Instead of commenting editorially on this matter, as Doctor Holzinger suggests, we have preferred to submit his criticism to Professor Kelley for his comment. This procedure commended itself to us for two reasons. First it meant less work for us; and second the result would be better. The reader may note, if he pleases, the order in which we state these considerations.

Professor Kelley replied almost immediately and we at once felt justified in resisting the temptation to comment editorially. In particular, we should not have added "homoscedastic" to the vocabularies of some of our readers.

We wonder how we ever managed to get along without this handy term. And its derivative, homoscedasticity, is even better. Do not these words fill a long-felt want?

Professor Kelley's letter follows:

In regard to the point raised by Mr. Holzinger let me say that the existence, or non-existence, of the equation $\Sigma_{1.t}=\sigma_{1.t}$ is an experimental matter and not a mathematical necessity, holding for all kinds of scatter diagrams or correlation surfaces. It is intimately connected with homoscedasticity, or equal variability of arrays. Galton

ran into the same problem and found that the arrays with which he dealt were substantially equally variable. In connection with a heredity problem he wrote:[1]

"I was certainly astonished to find the variability of the produce of the little seeds to be equal to that of the big ones; but so it was and I thankfully accept the fact, for had it been otherwise I cannot imagine from theoretical considerations how the typical problem could be solved." Galton's work led to the discovery of the normal correlation surface and in this arrays are equally, not proportionately, variable. I can think of no more sound assumption to make than that the correlation surface involving a single test score and a "true" test score constitutes a homoscedastic system[2]; but I will be the first to grant that it is an assumption, to be checked against experimental findings wherever possible. Should it not prove so we must regretfully accept the fact and exercise greater ingenuity than Galton imagined himself possessed of, to discover the lines of solution of the problem.

The old problem, reiterated by Mr. Holzinger, involves at least two things (a) the nature of the variability of successive arrays, and (b) the equivalence of the units in a test or scale. Until these are experimentally determined for a given test, I think we shall make fewer errors in interpretation if we assume that arrays are homoscedastic (equally variable) and that the units of measurement throughout the test are equal.

In closing, I would say that with reference to scales of the Binet type, which assume equivalence of successive age intervals, I think we already have abundant evidence to refute assumption (b). Accordingly, I do not believe the equation $\Sigma_{1.t}=\sigma_{1.t}$ holds for narrow and wide ranges in the case of the Binet scale. Not having two comparable forms of the Binet scale this matter is somewhat difficult to test, but a way for doing so can surely be devised.

Should the preceding observations tend to enlighten rather than confuse the readers of the JOURNAL I should be glad to have you print them in connection with Holzinger's criticism.

National Association of Directors of Educational Research

(E. J. ASHBAUGH, *Secretary and Editor*)

Perhaps nothing that can be reported in this issue of the Journal will be of greater interest to our membership than a statement concerning the meeting of the Executive Committee last month. Because of the removal of Dr. Buckingham from the University of Illinois to Ohio State University as noted in last month's items, certain questions arose as to the relation of the Association to its official organ. Final action of the Association cannot be given at this time, but it should be said that the Executive Committee after a thorough analysis of the situation agreed unanimously to stand for a journal of educational research whose board of editors and editorial policy should be guided by the Association.

A large portion of the time of the meeting of the committee was given to the tentative formation of the program for our next annual meeting. President Rugg informed us that he had been notified by the secretary of the Department of Superintepdence that our association is one of the TWELVE organizations invited to cooperate with the Superintendents at their Chicago meeting. This was encouraging news, showing that the great organization, which declared independence in control of its own meeting last winter, recognized the interest of its members in the work of our association and the real value of our programs. Certainly this year to an even greater degree than ever before, we may expect a hearing. We shall undoubtedly have an opportunity to impress the educational leaders with the importance of our work. The biggest and best program ever is the ambition of the president upon whom

falls the burden of securing the participants. We may have full confidence that he will realize his ambition.

Remember the place—CHICAGO.

The time—Last week in February.

The purpose—Selling the work of the National Association of Directors of Educational Research to all good school people.

You—Make your plans and reservations now.

The 1918–1920 Biennial Report of the State Department of Public Instruction of Wisconsin entitled "Educational Progress in Wisconsin" and edited by Mrs. C. W. Flemming has just come to your secretary. Mrs. Flemming, who is leaving the department to spend this year in Teachers College, Columbia University, leaves behind her a well-written report of work done and a valuable list of suggestions for the future work of the bureau of educational measurements in the department. Only a summary of the summary can be included here, but it is hoped that the department was able to send a copy to each of our members. The first paragraph will be quoted since it can scarcely be condensed. "The effort of the first two years of state supervision by means of educational measurement (1916–1918) was (1) to convince the school men of the state of the need of experimental study of school problems; (2) to familiarize them with the technic of the administration of tests; and (3) to train them in the interpretation and utilization of a few standard tests. Emphasis during the past biennium has, perhaps been given to the third objective."

Visitation, conferences, lectures, demonstrations, institutes, state-wide cooperative studies and preparation of special aids and reports have all been used to attain the desired ends. From the report it would seem that the tremendous amount of effort expended has been justified in part by the results accomplished and in part by the foundation laid for future activity.

Perhaps this is the proper place for the announcement of the change of address of your secretary. On invitation of Dr. Buckingham and the other proper authorities of Ohio State University to become Assistant Director of the new Bureau of Educational Research, Dr. Ashbaugh resigned his position at the State University of Iowa to accept this invitation. Dr. Ashbaugh was in charge of the Bureau of Educational Service in the Extension Division there for seven years where he gained a wealth of experience which will be directly applicable to the problems in his new position. During the first year of organization of the new bureau he will give a part of his time to teaching "School Administration" in the College of Education.

Announcement should also be made of change of position or address of other members.

Mr. P. C. Packer has resigned as assistant superintendent of schools in Detroit. He is teaching this first semester at his Alma Mater, State University of Iowa, and will spend the second semester in graduate study at Teachers College, Columbia.

Mr. W. W. Coxe of the Vocational Bureau, Cincinnati is spending this year in graduate study at Ohio State University.

Dr. Chas. Fordyce has resigned the deanship of Teachers College at the University of Nebraska to devote his entire time to the work of educational research in that institution.

Miss Henrietta V. Race, formerly Director of Intelligence Measurement in the Schools of Kansas City, Missouri, has become the Director of the Bureau of Psychological and Educational Measurement at Youngstown, Ohio.

Journal of Educational Research

OFFICIAL ORGAN OF THE NATIONAL ASSOCIATION OF DIRECTORS OF EDUCATIONAL RESEARCH

VOL. IV NOVEMBER, 1921 No. 4

CONTENTS

Entered as second-class matter at the Post Office at Menasha, Wisconsin, under the act of March 3, 1879.

Journal of Educational Research

Copyright 1921, by
Public School Publishing Co.

Official Organ of the National Association of Directors of Educational Research

PUBLISHERS' NOTICE

THE JOURNAL OF EDUCATIONAL RESEARCH is edited by Doctor B. R. Buckingham and a representative staff of associate and contributing editors. It is published by the Public School Publishing Company and is the official organ of the National Association of Directors of Educational Research. In the fullest sense it belongs to the subscribers and to those who contribute to its pages.

The agreement to publish a magazine for the Bureau of Educational Research of the University of Illinois has been annulled in accordance with its terms. In fulfillment of our obligations, assumed in that agreement to our subscribers, we now furnish this Journal published under the same editor and with a representative editorial staff.

But in establishing a Journal independent of any single institution, we wish to acknowledge a debt which we feel educational progress owes to the University of Illinois for its early recognition of the efforts which so many workers in educational research were ready to make in support of a journal in this field. The name of the magazine may be changed with the next volume, beginning with the January number, in order to give expression to its representative character. If our subscribers have a preference for the name, the publishers and the editor will be very glad to receive an expression of their desires.

JOURNAL *of* EDUCATIONAL RESEARCH

VOLUME IV NOVEMBER, 1921 NUMBER 4

THE CASE FOR THE LOW I. Q.

JOHN L. STENQUIST

Bureau of Reference, Research and Statistics, Department of Education, New York City

ILLUSTRIOUS SCHOOL FAILURES

Cases in which illustrious (not to include "merely successful") men and women were, while in school, diagnosed as failures by their teachers have often been cited. Many of the men and women, who later became world authorities in their fields, were called at best but mediocre. Linnæus' gymnasium teacher told his father that he was unfit for any profession. Yet this boy later was to revolutionize the science of botany.[1] Charles Darwin says in his autobiography that he "was considered by all his masters and by his father as a very ordinary boy, rather below the common standard of intellect." Napoleon Bonaparte in the final examination at his military school stood forty-second in his class. We may well ask with Swift, "who were the 41 above him?" Robert Fulton was called a dullard because his mind seemed filled with things outside of school. Priestly, the great chemist, had "an exceedingly imperfect education." Pasteur "was not at all remarkable at school. Books and study had little attraction for him." M. Pierre Curie, late professor of physics at the University of Paris with his wife co-discoverer of radium, "was so stupid in school that his parents removed him and placed him under a private tutor." Such a list as this could, if space permitted, be continued to great length. Many men who today are national or world figures, but who had a poor school record, could be cited.

Granting that these cases constitute but a minority—and granting also a certain tendency to exaggeration by biographers who love contrasts—these cases are still too numerous and important to be ignored. The fundamental fact remains that the

[1] Citations are from Swift: *Mind in the making*, chapter 1.

abilities of many pupils are widely misjudged in school, and abilities are either unperceived or misunderstood because of arrested development, poorly suited courses, stereotyped curricula, or general lack of sufficiently broad means for estimating ability.

No claim is here made that *all* so-called low I. Q.'s are misjudged—only that *many* are.

The Large Percent of Low Intelligence

That a great majority of pupils who enter the first grade drop out even before the end of the first year of high school is well known. Strayer's study of 318 cities quoted by Terman shows that of those who enter the first grade, on the average only 37 percent enter the first year of high school, 25 percent the second year of high school, 17 percent the third year, and 14 percent the fourth year. Studies by Ayres and Thorndike also show the same general tendency. Terman says "It is not uncommon for one-third to drop out without finishing the first year of high school." Retardation and elimination figures from every city annually offer additional testimony of the same general facts in elementary as well as in high schools. Terman believes that "not all of this elimination is traceable to inferior mental ability, but that a large part is due to this cause there is no longer room for doubt." With this general statement all will, of course, agree. The question, however, of just how much is due to actual lack of intelligence in its broadest sense, we do not know. Terman presents much evidence to show that with the use of general intelligence tests pupils who have low intelligence and who will drop out can be largely discovered beforehand.

But a situation in which over 80 percent of the pupil population is eliminated before reaching the goal is not greatly helped by the statement that most of the pupils who thus are eliminated haven't the general intelligence to proceed further. Is it not rather an indictment both of the curricula, *and* of the tests (which select too much on identical bases)? Terman suggests the query, "Are high school standards too high?" We might also ask are they too narrow? Or, in general, too far removed from the kinds of mental capacities of pupils? If such great numbers of the school population haven't the kind of ability we call general intelligence, why call it general?

What Is General Intelligence?

Certain it is that the term general intelligence is sorely in need of definition, for by the average person, and even by a large number of specialists in educational measurement, it is accepted at face value to mean just what it says. But is it not a loose use of terms that permits us to use the name "general" intelligence to designate mental traits which are painstakingly limited to the literary-academic tasks of our present intelligence tests? Are we not misleading when we say that he, and (in effect) *only* he has general intelligence, who, with paper and pencil, can effectively do such things as, solve simple problems in arithmetic, state the opposite for each of a list of words, insert a number of deleted words in sentences, arrange words in certain logical relationships, decide whether a given number or word is identical with another, write the seventh letter of the alphabet, arrange a jumble of words to form meaningful sentences, make a cross that "shall be in the circle but not in the triangle or square," state which day comes before Sunday, write whether a sentinel should be trustworthy, indicate whether alliteration is a form of pentameter, show whether cessation of belligerency is ever desirable, declare "what one should do if it is raining when we start to school," repeat "we are having a fine time. We found a little mouse in the trap," repeat "3–1–7–5–9," give the greatest possible number of words in one minute which rhyme with "day," or any combination of such tasks that may occupy the 30 to 45 minutes given to an average present day intelligence test!

What sort of mentality has the individual who makes a low score in such tasks but who when he drops out of school has the ability to organize a gang that is all but undissolvable? Or who drops out of school and builds up a world-wide business on the identical ground where "brighter" men have failed? Or who can wrest from a Robinson Crusoe situation a triumphant career? Or even he who can start a balking automobile abandoned by "superior" persons—men of higher I. Q.'s? Or what shall we say for the lamented low intelligence of the New York boy who escaped from an institution for mental defectives and who before the authorities recaptured him had obtained and was holding a job paying him $37 per week as a foreman in a blacksmith shop?

To say that there are but few such cases is untrue. Even through the illustrious cases do constitute but a minority, who

shall estimate how many more of that large percent who drop out of school because it is unsuited to their needs would develop careers of marked usefulness, if their real abilities were discovered and trained?

To say that such persons as those cited (except, perhaps, such cases as the last mentioned) are not possessed of general intelligence is to quarrel with words. Though they may classify as "low I. Q.'s" by present-day intelligence tests, surely we are on uncertain ground if we take such results at face value and consider the cases closed.

It is a question of what our tests measure, a question of what we mean to include under the term general intelligence.

If we examine the type of criteria by which nearly all these tests are justified, we find that they consist in the last analysis essentially of teachers' estimates of pupils' ability in school, plus records in other academic tests. But our major contention is precisely that for many children the teachers' estimates and their academic record is merely an estimate of success in bookish tasks, and that here it is that fallacies of intelligence ratings creep in.

It is submitted that these intelligence tests, at best, detect only those academic qualities of pupils which are noted by teachers, and which, it is freely granted, are of great importance for success in ordinary school curricula, but which do not constitute the whole of general intelligence. Of this our abler investigators[2] are fully aware, but the average giver of tests is not aware of it; or if so, he overlooks it.

Other Kinds of Intelligence

As a matter of fact, it seems clear that intelligence may be of many kinds. Thus, for example, the campaign manager exhibits a quality differing sharply from that of the locomotive engineer; while the kind of intelligence required to lay out the construction work of a Woolworth Building is not very like that needed to write a forceful letter, and this in turn is not very like that employed in painting a great picture, or inventing a great machine such as the modern linotype.

[2] Thorndike, E. L., "Tests of intelligence, reliability, significance, etc.," *School and Society*, v. 9, February 15, 1919; Henmon, V. A. C. "Measurement of intelligence, *School and Society* v. 13, February 5, 1921.

While it may be true that a certain minimum body of "common sense" mental ability, and *some* general academic information underly all such activities, we know from at least a few correlations obtained (one of which appears later) that the relationship is not very close—though it is, to be sure, positive.

If we had trustworthy criteria of ability in social leadership and in the various political and mechanical arts and sciences, it might be possible to devise intelligence tests that would be more nearly "general" than those we now have. This, however, is a more difficult matter than to devise tests of academic ability. Again, while to measure in this wide sense the present general intelligence of our school population represents a heavy task, to prognosticate its *potential* ability would be a truly Herculean undertaking. But this is not equivalent to saying that it can't be done. Much of the same methodology and technic which we already have would apply, and progress in this direction may be looked for. Current literature is already sprinkled with discussions of the limitations of what our present so-called general intelligence tests measure.

General Intelligence and Mechanical Ability—Results of Tests

The tests of mechanical ability herein described may serve as an example and case in point, showing another type of intelligence and also emphasizing the need for clearer definition of just what we mean when we say a child has but little general intelligence.

During 1919–1920 several hundred boys in a New York City public school (P. S. 64, Manhattan) were given a very exhaustive intelligence rating by means of the combined results in the following well-known tests:[3]

A. The Intelligence Tests

National Intelligence Test A and B

- Haggerty Intelligence Test, Delta 2
- Otis Intelligence Test
- Myers Mental Measure
- Thorndike's Visual Vocabulary

[3] Stenquist, J. L., "Better Grading Through Use of Standard Mental Tests," Bureau of Reference, Research and Statistics, Board of Education, New York City.

The results of these six tests were pooled, giving equal weight to each, and the final rating was called the composite intelligence score. These boys were next given a series of mechanical tests, consisting of the following:

B. THE STENQUIST MECHANICAL TESTS

Assembling Series I: Ten unassembled commercial mechanical articles in a 5 x 2½ x 24 inch box, divided into ten compartments, to be assembled; these include a cupboard catch, an ordinary clothes pin, a three-piece paper clamp, six links of ordinary safety chain, a simple bicycle bell, a shut-off clamp (used on rubber tubing), an ordinary wire rubber stopper (such as is used on pop bottles), a push button, a simple three-piece lock, and a mouse trap.

Assembling Series II: Ten more articles like the above, consisting of a four-piece elbow catch (used on almost every cupboard door), a rope coupling, a toy pistol, an expansion nut, a window sash fastener (such as is commonly found on windows), a simple pair of calipers, an expansion four-piece rubber stopper for bottles, a four-piece paper clamp or clip, a double hinge (such as is found on every screen), and a simple lock.

Picture Test I: Consisting of 78 picture-matching problems in which the pupil is required to determine which one of five pictures belongs with, or is a part of, one key picture. The pictures treat general mechanical subjects such as mechanisms, tools, toys, machines, and their parts.

Picture Test II: This is similar to Picture Test I, but involves some language. Sixty questions are included referring to numbered parts of machine. These call for a certain type of mechanical reasoning, requiring the ability to think in terms of mechanical problems. Seventeen questions are devoted to matching pictured parts of mechanical toys.[4]

These mechanical tests inter-correlate on the average between 0.6 and 0.7 One test of actual manipulation of objects, such as Series I, correlates about as high with either of the picture

[4] Stenquist, J. L., *Measurements of Mechanical Ability*.

tests as it does with a second series of models to assemble. On the whole, any one of the four tests affords an important indication of a general ability that may for convenience be called "general mechanical aptitude"—general in the sense that it does not pertain to any special trade, and mechanical, as is more or less obvious, from its nature.

Comparison of Results

If we now compare the results in the two types of examination we may observe the following points:

As to the correlation between the Assembling Test, Series I, and the composite intelligence score, $r = 0.23 \pm 0.04$, for 267 seventh- and eighth-grade boys. Between Assembling Test, Series II, and the composite intelligence score $r = 0.34 \pm 0.06$ for 100 seventh- and eighth-grade boys. Between Picture Test II and the same intelligence rating $r = 0.34 \pm 0.03$ for 296 seventh- and eighth-grade boys.[5]

If we now combine all of the four mechanical tests into one average T-Score,[6] and correlate it with the same intelligence rating, we find r drops to 0.21 ± 0.04 for 275 seventh- and eighth-grade boys.

The important inference to draw from these results is not with regard to the exact coefficients obtained, but with regard to the general fact of low correlation between the two kinds of ability here represented. Results obtained in the army for over 14,000 men bear out the same general fact.

Figure 1, in which each dot represents an individual, shows graphically the same fact.

An individual's position in general intelligence is thus shown to be largely independent of his position in General Mechanical Ability and Aptitude.

The Trustworthiness of the Measurements

As regards the reliability of our measure of general intelligence, comprised as it is of six excellent tests, any *one* of which would

[5] The corresponding r-value for Picture Test I was computed for a presumably greater spread of talent—namely for a group including sixth-grade boys as well as those of the seventh and eighth grades. The resulting value of r (which was 0.52) is not therefore comparable with the figures given in the text.

[6] T-score is the S.D. equivalent for distribution of twelve-year-old boys, with zero considered at -5 S.D. See Wm. A. McCall *How to measure education*, "A uniform method of scale construction," *Teachers College Record*, January, 1921.

generally be accepted as a measure of general intelligence, it constitutes an unimpeachable estimate of that type of ability which we now call general intelligence. In mechanical ability we have repeated tests of each of two types of mechanical tasks—the assembling tests involving skill, and the picture tests, involving mechanical information and reasoning, i.e., we have in fact four distinct measurements of each pupil. The reliability of our measures is, therefore, acceptable, and much better than is generally obtained.

The Validity of the Measurements

The validity of a test deals with the question of what it is that it measures—i.e., with correlations with criteria.

The question of what the intelligence tests measure has already been dealt with. As to what the mechanical tests measure we may first cite the correlations which have been found in comparing mechanical test scores with pupils' ranks in shop courses and in general science courses. The following instances are cited for the Assembling Tests and shop work:

27 boys seventh and eighth grade,	between test and shop rank,	$r = 0.83$
15 boys eighth grade	" " " " "	$r = 0.80$
24 boys eighth grade	" " " " "	$r = 0.42$
14 boys sixth and seventh grade	" " " " "	$r = 0.81$
18 boys sixth grade	" " " " "	$r = 0.90$
17 boys sixth grade	" " " " "	$r = 0.88$

The following correlations were found for the Picture Tests and shop work:

27 seventh and eighth grade boys,	between test and shop rank,	$r = 0.83$
53 high school	" " " " " "	$r = 0.53$
14 sixth and seventh grade	" " " " " "	$r = 0.51$
18 sixth grade	" " " " " "	$r = 0.59$
17 sixth grade	" " " " " "	$r = 0.59$
27 seventh and eighth grade	" " " " " "	$r = 0.84$

Correlations of the same order of magnitude were found between estimates of teachers of general science and scores in the mechanical tests.

These correlations are all subject to chance errors *which reduce them*. The true correlations are therefore higher.

Shop teachers' ranks are, of course, no better than regular teachers' ranks which have been criticized in the previous section.

But there is every reason to believe them *equally good*, and the same may be said of the judgments of science teachers. Were other and better criteria available these estimates would be excluded. In several of the above instances the average rank given by two shop or science teachers (intercorrelating 0.88 or better) were used.

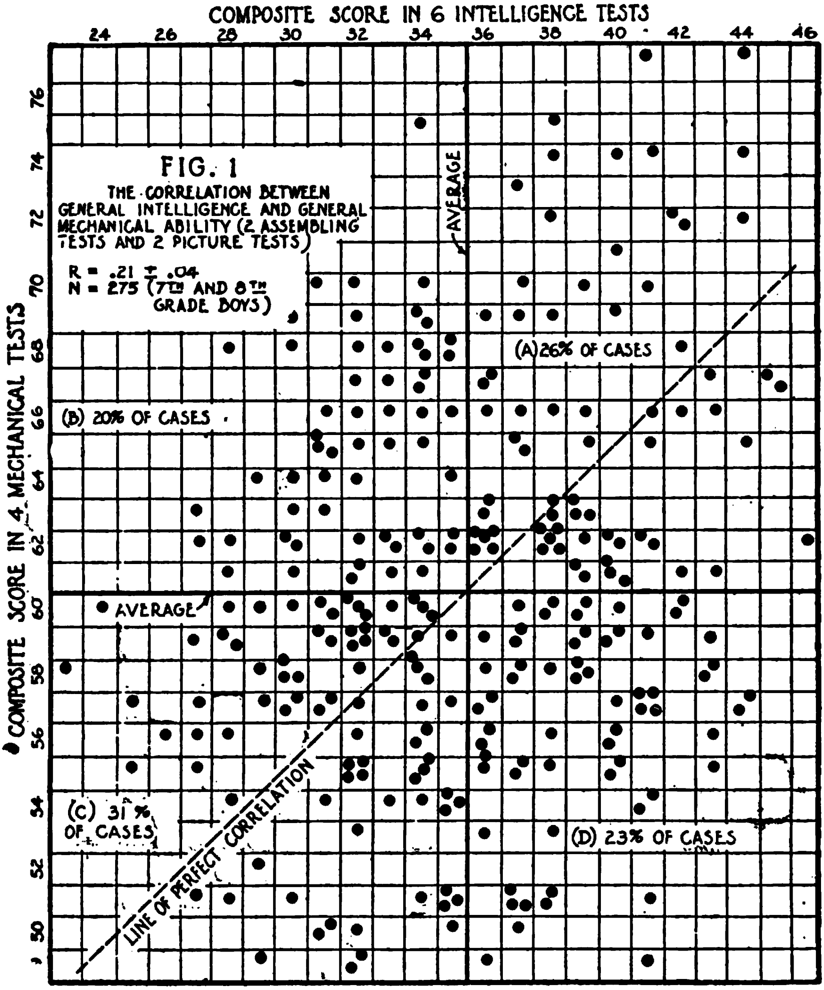
The mechanical tests may, therefore, be judged from these figures to detect to a marked degree the same qualities in pupils that are considered by shop and science teachers in judging pupils' relative abilities.

What is it that causes a pupil to stand out in this type of work? May it not be another type of intelligence that might well be called of *general* importance?

The second way of deciding what these mechanical tests measure is the very direct one of merely looking at the tests and judging what type of task it is that has been set up. Thus we may note at once that they represent an attempt (in all except Picture Test II) to get away from words. They deal with concrete and real things, as against descriptions of things. In the Assembling Tests opportunity is given to work with hands and mind, rather than to perform with a pencil only, or to juggle mental abstractions.

Analysis of Total Distribution

Figure 1 represents by dots each of 275 seventh- and eighth-grade pupils distributed according to their scores in the intelligence tests and in the mechanical tests. For convenience the figure has been divided into four quadrants each identified by a letter. The percent of the total number of pupils who fall in each quadrant is also indicated. Of the pupils in Groups B and C (all of whom are *below* average in general intelligence) two-fifths, or 20 percent of the total number, are in Group B—i.e. are *above* average ability in the mechanical tests. The pupils in groups A and B combined, or 46 percent, are above average in mechanical ability. Of these, Group A contains more than one-half (or 26 percent of the total number). These are also above average in intelligence. But for the fact that our tests show the marked ability of such pupils in mechanical ways, it is unlikely that many of Group A would be encouraged to look toward careers in mechanical fields, since they have marked abstract intelligence.



Conversely, without mechanical tests, those in Group D would not be known to be deficient in mechanical ability. Since they are above average in intelligence their general superiority might be, and no doubt often is, falsely inferred. Considering mechanical ability alone we may say that Groups A and B would likely succeed in this direction, while Groups C and D would not be likely to do so.

Again, if we were to rely merely on the intelligence tests, all pupils in Group B (one-fifth of the entire number under

consideration) would fail to be recognized as having a highly useful other ability, and as having it to a marked degree. Consider next Group C which consisted of pupils who are low in both tests. It is not without value to have this double negative of information. At least advice can be given less blindly than without such information. Again, there may be quite different types of abilities in which some of these may excel. Having them segregated we can proceed more intelligently than otherwise, to say the least. Less progress should be looked for, for one thing.

In short, the mechanical tests have given us important clues as to abilities which would not be revealed by the abstract intelligence tests alone. Though the correlation is positive it is so low as to permit wide differences in deviation. These are measures of abilities untouched by so-called general intelligence tests.

It may be thought, however, that the mixture of abilities revealed by combining the assembling and picture tests is less illuminating than would be the abilities revealed by either type of test taken alone. In order to examine this question the records of 267 seventh- and eighth-grade boys in the assembling tests were separately plotted in relation to their scores in the intelligence tests. The correlation with "general intelligence" (0.23) was practically the same as in Figure 1, and the percents in each quadrant were either identical or different by but one point. The records, however, of 296 boys of the same grades in Picture Test II showed a somewhat closer correlation (0.34) with scores in intelligence. Yet the percent of pupils in each quadrant was highly significant. In A there were 28 percent of the pupils, in B 22 percent, in C 36 percent, and in D 14 percent. The reader's attention is especially directed to the fact that Group B contained 22 percent of all the pupils—in other words to the fact that more than one-fifth of them were low in intelligence but high in the ability measured by the picture test.

The Relative Importance of These Two Kinds of Ability

Of the relative importance of each of these two types of ability, readers must form their own conclusions. But it should be kept in mind that we are living in a world that is dominated on every hand by every known form of mechanical device and

machine. Every moment of present-day life is influenced directly or indirectly by the products of mechanical skill and genius. Is it not important that ability in this field should be discovered and developed? Rather than merely to dismiss our apparently stupid pupils as low in "general intelligence," and to relegate them to some convenient class, our time might profitably be spent in disclosing other kinds of intelligence of which they may be possessed.

Possibly it would be more appropriate to designate these mechanical tests by some other name for they are mechanical only in a limited sense. On the mental side they call for the ability to recognize parts of ordinary mechanical devices, for the ability to make judgments as to the reasons for the particular size, shape, weight, and nature of the parts—in short, for the mental ability to think through in some degree the same steps as those employed by the designer of each machine. Manually, they call for the dexterity required to put parts together to form the completed machine or device after it has been decided how they should go. Much of the performance of a typical child is, of course, mere trial and error manipulation, in which he hopes somehow "to make the thing work." But the nature of the various models is such that only a very low score is possible for the individual who depends merely upon thoughtless manipulation of the parts. A generous amount of the best kind of thinking is thus required to make a high score. It involves accurate perception, reasoning, and judgment applied to each model. In so far, therefore, as these mental processes are of general importance in every-day life, the ability demonstrated in assembling these models perfectly could well be called general intelligence.

Fictitious Stigmas

There is a strong and universal notion that a low score in such tasks as have here been called intelligence tests constitutes a disgrace, that must be shunned at all costs. To fail to receive a high rating in intelligence is regarded as a calamity. This feeling has come about partly through the loose use of the term general intelligence, and partly through a distorted estimate of the rôle of intelligence in human conduct. Absurd as it may seem, there is a brief, and a reasonable one, which can be held for the I. Q. which is actually low—as well as for the supposedly low

I. Q. For just as in man we find enormous individual differences in intelligence, so (fortunately) in the work of the world we find equally great variation in the character of the work to be done. As a matter of fact, the outstanding industrial tendency of the past decade has been to reduce the number of skilled jobs and increase the number of unskilled ones. The constant tendency of our modern machine age is in this direction, be it right or wrong. Again, consider the hundreds of thousands of menial tasks outside of industry that somebody must perform in every society. Is it not clear that happiness, contentment, and efficiency in such jobs are far more apt to come with a *low* I. Q. than with one that is high? Indeed, even when we consider the world's sweetest and most lovable characters, it is not always their high general abstract intelligence that makes the strong appeal. Haven't we in the academic atmosphere of our schoolrooms come to value the intellectual side of human nature out of proportion to its real significance in life? Surely far worse calamities can befall the human animal than that of being pronounced of low intelligence. Physical disease, a crippled body, an insane or actually feeble mind, with the multitude of tragic afflictions which this may imply—these and many other lamentable conditions which may befall should be kept in the background of our mind when we feel inclined to bemoan the lot of the stupid individual.

SUMMARY

This account attempts to point out some of the fallacies that are prevalent in the present-day conception of intelligence tests. It recalls the many cases of illustrious men who were called school failures, and calls attention to the large percent of pupils who at present appear to lack sufficient mentality to carry on current curricula, and suggests the query "Is it the curricula or the mental ability of the population that is at fault?" It criticizes present-day intelligence tests as narrow and academic in scope, being based largely on school success; suggests the loose use to which the term "general intelligence" is often put; and maintains that there are in fact many other kinds of intelligence than are now being measured by tests of that name. As an illustration the results of a study of mechanical ability are offered. Here it is shown that at least 20 percent of the pupils from a

typical school, who are below average in general abstract intelligence, are *above* average in the kind of ability required in four mechanical tests, the detailed nature of which is described. It is submitted that such ability may be of quite as general importance as that required to score high in the abstract general intelligence tests, in view of the fact that present environment is so largely permeated with the fruits of mechanical genius and applied science. Finally, it is maintained that there is a strong, but wrong tendency to attach a stigma to pupils scoring low in so-called intelligence tests Even for pupils whose true general intelligence is found after adequate testing to be low, there is ample opportunity for useful and happy lives—lives concerned with tasks for which they are, in fact, better adapted than are individuals of high intelligence.

MEASURING THE EFFICIENCY OF TEACHERS BY STANDARDIZED TESTS[1]

SAMUEL S. BROOKS
District Superintendent, Winchester, N. H.

My last article told how we used standardized tests and scales to measure the progress of pupils and to tell when they were ready for promotion. This one is planned to show how, at the same time, we were measuring the ability of the teachers to get results.

Besides knowledge of subject matter, one may recognize five main factors in a teacher's efficiency: (1) managing ability; (2) natural aptitude for the work; (3) method and technic of teaching; (4) interest and industry in her work; (5) and that vague thing, personality, somewhat indefinable but generally admitted to include character, temperament, personal appearance, manners, tact, etc. She must demonstrate ability to organize and manage a school in an orderly manner before any of her other abilities can do their work. With all the other factors present, a teacher's success can be but mediocre if she lacks greatly in natural ability as applied to teaching. She may have all the other virtues but if she lacks enthusiasm and industry she cannot inspire her pupils; and without an efficient method her other qualities will be ineffective. Finally, her personal qualities, ideals, and conduct must be worthy of emulation if she expects to influence properly the social and moral life of her pupils.

Now, no one of these factors can be accurately and objectively measured independently of all the others. But they all function cooperatively in getting results—results which are manifested in the development of knowledge, skill, and ideals among pupils. And these results *can* be measured by means of standardized tests.

Is it not customary to measure the efficiency of the workman, professional or otherwise, by the amount and quality of the work he turns out? The efficiency of the wood chopper is gauged by the number of cords of wood he can chop in a definite length of time; of a bricklayer, by the number of bricks he can lay in a day; of

[1] This is the sixth article by Superintendent Brooks on the general topic, "Putting Standardized Tests to Practical Use in Rural Schools."

the farmer by the per acre yield and profit of his crops; of the lawyer by the percent of cases he wins for his clients; of the doctor, by the proportion of cases he cures; and so on, for almost any line of human endeavor we could mention. Experience has set certain standards of achievement in every kind of work and the efficiency of the worker is judged by the ratio of his product to these standards. If he does only three-fourths as much as the standard he is only 75 percent efficient.

Then why should not the efficiency of teachers be measured by the amount of work they turn out? Too long has efficiency been taken for granted or, at best left to the judgment of supervisors making guesses based on classroom observation more or less perfunctory of teachers' good looks, engaging personality, show of energy and enthusiasm, evidence of preparation, handling of supervisor's pet methods, etc. Although such observation is not without value in helping secure a fair estimate of a teacher's ability it does not alone furnish a safe and sane basis for judgment, and any teacher so judged to be inefficient has a right to complain of unfairness of treatment. Judgments based on mere classroom observation are not fair either to the teachers or to the taxpayers. The reasons why this is so were summarized in my first article, but they will bear repeating here.

(1) Such observations do not furnish a sound basis for judgment; (2) the superintendent's opinions are quite apt to be colored by personal prejudices toward an individual teacher or her methods; (3) classes often show at their worst in the presence of visitors; (4) even the teacher may fail to do herself justice under the critical eye of the superintendent; and (5) classroom observation takes no account of the actual results the teacher may be getting. Furthermore such observation is not only unfair but inaccurate. It is inaccurate because of all the reasons just given and because (a) some teachers do excellent work when the superintendent is present and shirk all the rest of the time, and (b), if such teachers do their own testing, even the results may be made falsely to appear satisfactory.

If the education of a child consists in his acquiring certain knowledge, skills, habits, and ideals that will make him a useful and desirable member of the society in which he lives, and if teaching is the proper leading and directing of the child in utilizing his natural abilities to acquire these things with the least possible

expenditure of time and energy, then why is it not eminently fair to all concerned to gauge the teacher's efficiency by measuring at definite intervals the progress her pupils are making in the acquisition of the prescribed knowledge, skills, habits, and ideals, provided we have well-defined standards of achievement for each grade such as the standardized tests furnish?

Anyway, I put the question squarely up to the teachers of my district at one of the teachers' meetings held early in the year. They were asked to decide whether they would prefer to have the superintendent estimate their teaching efficiency on the basis of what classroom observation he could make in schools so widely scattered, or according to the progress made by their pupils as measured by standardized tests.

As I had expected, the question evoked a lively discussion and some well-founded objections were raised. Most of the teachers were ready to admit the inaccuracy and unfairness of ordinary methods of rating teachers, but insisted that there was a large probability of the same weaknesses in the plan I proposed. Their chief objections were: (1) that knowing they would be judged by the results of the tests some teachers would be tempted to cheat in giving the tests, thereby perhaps gaining a higher rating than better and more conscientious teachers who gave the tests honestly; (2) that since there are in most schools a sprinkling of mentally deficient or even feebleminded children who under the most efficient teacher cannot be expected to make normal progress, the records of such pupils when averaged with those of normal children would seriously and unjustly lower the rating of the teachers; and (3) that of two teachers of equal ability one might have a school whose pupils averaged so much higher in intelligence than those of the other that she would undeservedly obtain a much higher rating. The majority thought that, if these principal objections could be satisfactorily disposed of, the plan would be worth trying. The few teachers who displayed marked lack of interest in the subject had already on other grounds shown themselves to be of the time-serving variety. I therefore ignored their attitude. But I wanted the intelligent acquiescence of the better teachers in some sort of a reliable teacher-rating scheme.

The first two objections I had foreseen and prepared for. As to the first, I explained that most of the tests were furnished in

two or three different forms so that the same forms would not have to be given twice in the same year. This would obviate the possibility of any teacher drilling pupils on the exact contents of a test, drill along the general lines of work suggested by the tests being not only legitimate but desirable. Furthermore, I pointed out that my plan of checking the work of the teachers in giving the tests would insure the immediate discovery of any serious attempt at cheating on the part of dishonestly inclined teachers—such as allowing more than the allotted time for each test or giving illegitimate aid to the pupils during the tests. This plan was for me to repeat in each school one or two of the tests *after* the teachers had given them all. Then if there was any great discrepancy between the results of the tests I had given and those a teacher had given, such discrepancy would indicate either dishonesty or gross carelessness in giving the tests.

The second objection offered a good opportunity for a discussion of intelligence tests and their uses. I passed around some samples of the Otis Group Intelligence Test and explained how, by the use of such tests, we could locate the pupils who were mentally incapable of making normal progress. The progress records of these pupils could be thrown out in calculating the teachers' ratings, and we might use only the records of pupils who graded 80 percent of normal or better by the intelligence tests.

The third objection was one which had not before occurred to me. I suggested that we leave the matter until our next meeting by which time I hoped to have a satisfactory solution. The plan I finally worked out and which was accepted as satisfactory by the teachers follows: From the results of the June tests the average scores by grades for each test were to be calculated for each school. Each of these average grade scores was to be divided by the corresponding standard score thus giving the percent which each grade score was of normal.

Table I illustrates the method by which these percents were obtained. The figures opposite the pupils' numbers are the rate and comprehension scores in reading for a fifth grade in the June tests.

All percents similarly derived for each school to be averaged to give the teacher's percentage mark. Then, to offset the differences in intelligence between schools, if the average of the I. Q.'s

in a school was less than 100, the difference between it and 100 was to be added to the teacher's mark and if the average of the I.Q.'s was more than 100, the difference was to be subtracted from the teacher's mark. This procedure served in the one case to discount the part of a school's progress that was due to superior native intelligence and in the other case to give the teacher an allowance to offset her school's mental disabilities. This plan disposed of the third objection mentioned above. Its accuracy of course depends in large part on the degree of correlation between the scores in intelligence tests and the scores in achievement tests. That the correlation is high will be shown later in another article. This scheme does away with the necessity of discarding the scores of subnormal children in calculating the ratings of teachers, although such discarding would save considerable work without materially affecting results.

Below are given concrete illustrations of how the ratings of several teachers were obtained at the end of the year. The first two are both very competent and successful teachers. In A's school the average of the I.Q.'s of all the pupils was 88. This school had thirty-two pupils, four of whom graded as feeble-minded. Twenty of them had I.Q.'s of less than 90. Only five

TABLE I. JUNE SCORES OF A FIFTH GRADE IN READING

Pupil Number	Rate	Comprehension
8	108	26
2	98	20
3	73	14
4	85	19
5	101	25
6	95	21
7	50	8
8	105	20
Average	89.4	19.1
Standard	93	20
Percent average score is of standard	96.1	95.5

had I.Q.'s over 100 and the highest was 122. In B's school consisting of thirty pupils, the average of the I.Q.'s was 111. The intelligence level in this school was unusually high just as in the other it was unusually low. There were no feebleminded children and only one pupil graded as very dull. Eighty-three percent of the pupils were normal or above. Three had I.Q.'s above 140. (All intelligence tests were given, corrected, and scored by the superintendent.)

Table II gives the grade percents (computed as shown in Table I) on each test in A's school—also the general average for the whole school. The 78, for instance, at the top of the second-grade column in Table II means that the second-grade average score in rate of silent reading was 78 percent of the second-grade standard score. In comprehension of silent reading their average score was only 69 percent of the standard score, and so on for each subject and for each grade. There are 98 of these percents in the

TABLE II. GRADE PERCENTS ON EACH TEST. TEACHER A

Subjects	Grades						
	II	III	IV	V	VI	VII	VIII
Silent Reading							
Rate	78	81	86	90	93	99	98
Comprehension	69	77	80	80	78	85	84
Addition	85	87	92	92	94	96	98
Subtraction	93	91	93	92	93	98	98
Multiplication	93	92	97	93	98	98	100
Division	74	79	82	87	92	95	97
Mixed Fundamentals	72	76	77	80	82	91	93
Arithmetical Reasoning	......	70	73	78	81	88	87
Spelling	82	80	80	77	83	86	87
Writing							
Speed	96	102	101	104	111	107	109
Quality	67	68	67	65	62	62	65
English Organization	......	92	94	92	88	94	93
Visual Vocabulary	......	83	84	81	77	86	89
Geography	......	......	82	80	85	89	92
History	......	......	78	72	82	84	90
Grade Averages	80.9	82.9	84.4	84.2	86.6	90.5	92.0

General Average = 86.3

table. The general average for the school was obtained by adding all of them and dividing the sum by 98. The general average in this school was 86.3 percent, which means that the average progress of the school, as measured by the standardized tests, was 86.3 percent of normal. Table III gives the same data for B's school. In this case the general average was 108.4 percent of normal.

TABLE III. GRADE PERCENTS ON EACH TEST. TEACHER B

Subjects	Grades						
	II	III	IV	V	VI	VII	VIII
Reading							
Rate	102	105	111	114	117	123	122
Comprehension	93	101	104	104	102	109	108
Addition	107	109	114	114	116	118	120
Subtraction	114	112	114	113	114	119	119
Multiplication	114	113	118	114	119	119	121
Division	96	101	104	109	114	117	119
Mixed Fundamentals	94	98	99	102	104	113	115
Arithmetical Reasoning	......	92	95	99	104	110	109
Spelling	104	102	102	99	105	109	108
Writing							
Speed	118	124	123	126	123	129	131
Quality	89	90	89	87	84	84	86
English Organization	......	114	116	114	116	116	113
Visual Vocabulary	......	106	103	108	108	105	111
Geography	......	......	104	102	111	111	114
History	......	......	96	95	106	107	109
Grade Averages	103.1	105.1	106.1	106.7	109.5	112.6	113.7

General Average = 108.4

Then according to the rating plan described above: A's rating = General average for A's school + (100 − Av. I.Q.) = 86.3 + (100 − 88) = 98.3. And B's rating = General Average for B's school − (Av. I.Q. − 100) = 108.4 − (111 − 100) = 97.4.

These are the ratings of two teachers of undoubted ability but with schools widely varying in average intelligence and rate of progress. Yet the rating shows the teachers to be of about equal ability. The difference in progress in the two schools is due to difference in the average mentality of the pupils. It would be

eminently unfair to expect equal progress with the two schools or to rate A as a poorer teacher than B because progress in A's school was less than progress in B's school.

Now let us consider the cases of two teachers of widely different ability but with schools approximately equal in size and in the average intelligence of pupils. Teacher C is a normal-school graduate, with several years' experience but with apparently little aptitude for or interest in the work, who tries to teach as she was taught regardless of her professional training. Teacher D is an enthusiastic girl of twenty years who has had one summer term at normal school and one year's experience. Apparently she got more out of her summer session than many do out of the whole course. Moreover, she has the ability to adapt her knowledge to classroom use.

Tables IV and V give the same data as Tables II and III but for the schools of C and D respectively.

TABLE IV. GRADE PERCENTS ON EACH TEST. TEACHER C

Subjects	Grades				
	II	III	IV	VI	VII
Reading					
Rate	80	84	88	92	95
Comprehension	75	81	80	85	90
Addition	84	85	90	91	94
Subtraction	92	90	89	96	95
Multiplication	90	89	94	90	96
Division	78	85	80	82	88
Mixed Fundamentals	73	74	75	78	84
Arithmetical Reasoning	......	70	74	79	82
Spelling	78	75	80	81	89
Writing					
Speed	87	92	93	98	101
Quality	67	70	65	65	58
English Organization	......	90	90	95	95
Visual Vocabulary	......	85	82	79	76
Geography	......	......	83	87	90
History	......	......	76	82	80
Grade averages	80.4	82.3	82.6	85.3	87.5

General average = 83.9

TABLE V. GRADE PERCENTS ON EACH TEST. TEACHER D

Subjects	Grades				
	II	IV	VI	VII	VIII
Reading					
Rate	95	98	103	107	110
Comprehension	86	94	97	97	95
Addition	100	102	107	107	109
Subtraction	107	105	107	106	107
Multiplication	107	106	111	107	112
Division	89	94	97	102	107
Mixed Fundamentals	87	91	92	95	97
Arithmetical Reasoning	......	85	88	92	97
Spelling	97	95	95	92	98
Writing					
Speed	111	117	116	119	116
Quality	82	83	82	80	77
English Organization	......	109	107	107	109
Visual Vocabulary	......	98	99	96	101
Geography	......	97	95	104	103
History	......	93	87	99	95
Grade averages	96.1	97.8	98.9	100.7	102.2

General average = 99.3

The average of I. Q.'s for C's school was 98.8, and that for D's school was 102.2. This is a slight advantage for D's school but not nearly enough to account for the difference in attainment in the two schools. Calculated as before,

C's rating = 83.9 + (100 − 98.8) = 85.1.

And D's rating = 99.3 − (102.2 − 100) = 97.1.

Here again the relative efficiency of the teachers is reflected in respective ratings of their schools when full cognizance is taken of the average intelligence in the two schools.

We use averages rather than medians in computing the ratings of teachers because the schools are small with few pupils in a grade. In larger schools with twenty or more pupils to a grade the median scores could be as well used in figuring grade percents. In such case one should not neglect to use median I. Q.'s as well as median scores. And it might be well to mention here that when the scores of subnormal children are thrown out of the reckoning

their I.Q.'s should be discarded also; otherwise the teacher's rating would be considerably affected.

Although, of course, this rating does not include everything that should be taken into account in estimating a teacher's worth to the school and to the community it nevertheless covers one of the most important factors to be considered and furnishes a fairly objective test by means of which on occasion a teacher can be convinced of her own inefficiency. Certainly if a teacher fails seriously in this phase of her work she can not profitably be kept on the payroll for the sake of her personal appearance, good moral influence, managing ability or any other factor or factors that go to make up a good teacher.

In addition to a substantial general raise in salaries throughout the district for the current year, the school boards were persuaded to grant special increases of one or two dollars per week to certain teachers who rated 95 percent or better with ratings calculated as described. None of the teachers who failed to get such a raise made any complaint of favoritism, nor could they consistently do so since they had themselves accepted the basis on which their ratings were determined. Furthermore the teachers are working this year with the understanding that they will receive bonuses at the end of the year of five dollars for every whole unit that they increase their ratings over those of last year, the bonus not to exceed fifty dollars. Thus, if a teacher's rating last June was 89.2 and next June it has increased to 94.4 she will have increased her rating $94.4-89.2=5.2$ or five whole units. Hence she will receive a bonus of twenty-five dollars. I know that most of the teachers are working hard for a bonus.

A CYCLE-OMNIBUS INTELLIGENCE TEST FOR COLLEGE STUDENTS

L. L. THURSTONE
Carnegie Institute of Technology, Pittsburgh, Pennsylvania

Intelligence tests were until recently used primarily to determine the mental age of children by the comparison of the chronological age with the mental age. When mental tests are given to adults, this type of analysis is no longer suitable since the intelligence of adults does not increase noticeably with age. The intelligence tests are used with adults in order to place a given person's intelligence with reference to the distribution of general adult intelligence. The scores are, therefore, more suitably expressed in terms of ranks or percentile ranks or of the standard deviation of the distribution of adult intelligence.

An intelligence test should be so adapted as to its difficulty that it will differentiate into significant and distinguishable groups the population for which it is intended. This cannot well be done if the test is either too difficult or too easy. In general, however, a fairly symmetrical distribution of test scores will be obtained for any given population with tests of considerable variation in difficulty. The test for which I am here submitting some norms was arranged in difficulty as nearly as possible to suit college freshmen.

During the last five years I have been experimenting with over fifty different varieties of mental tests with college freshmen at Carnegie Institute of Technology in its various departments. My procedure was the general one of giving the tests as early as possible in the freshman year, filing the scores until scholarship records and instructors' estimates could be obtained, and then determining the predictive value of each test by the correlations between its scores and the objective criteria of the students' abilities. I have found that instructors' estimates, on account of their unreliability, are, in general, unsuitable as a criterion by which to judge the predictive value of a mental test. The several instructors' estimates for the same student vary considerably more than the corresponding scholarship grades.

Lately I have used scholarship records almost exclusively as the general criterion to determine the predictive value of a mental test. The procedure is quite different, however, in work-

ing with individual tests, in which case one can use as a criterion the estimates of instructors because these estimates are then open to qualitative interpretation in the light of the examiner's personal acquaintance with the students tested. For the purpose of group testing of intelligence, the scholarship grades are perhaps the best available objective criterion.

There is considerable objection to the use of scholarship records as a criterion for determining the diagnostic value of a mental test; and this is not without justification, because the mental test may be a better index of intelligence than the scholarship grades. However, we must here sharply differentiate between the scientific problem of defining and diagnosing intelligence and the administrative problem of determining the relative predictive value of mental tests, entrance examinations, high-school certificates, interviews, and the like, for educational guidance. It is to the more restricted and immediate administrative problem that most American mental test investigators are devoting their attention. Let us hope that we may be able to tease out some scientific principles regarding intelligence and mental tests from our many laboriously tabulated forms of performance and correlation coefficients.

My justification for using scholarship grades instead of estimates of intelligence as a criterion for mental tests for college freshmen is that the student's retention and promotion in college depends more on his scholarship records than on any other single factor. The administrative question is this: By giving intelligence tests to our entering students, are we more able to predict their ability to do college work than by our present methods of admission? Available objective evidence, especially if we consider the combination of intelligence tests with evidence of high-school preparation in the fundamental subjects, certainly gives the affirmative answer to this question. Since scholarship is the one outstanding factor which determines academic success it is the only logical criterion by which to judge the predictive value of mental tests. It may be argued that the scholarship records do not predict success in life and that therefore some other criterion should be adopted for the selection of the students who are to be given a college education That is an entirely different question. The immediate question is to determine how intelligence tests can be of use in colleges constituted as they are at present. If

mental tests should be found to have no correlation with scholarship they could not be of use in admitting students, because they would not reduce the freshman mortality, even though they measured intelligence with absolute certainty. If the declared purpose of the colleges is considerably changed, and if the college curricula are fundamentally changed, it may be necessary for us to modify our mental tests so as to have predictive value in a different order of things. It is, therefore, quite possible that a mental test which is serviceable for admitting students to a college course may or may not be a test of intelligence. The empirical justification for mental tests is that they work, and the concept of intelligence which the tests are tacitly assumed to measure is after all an abstraction.

The criterion of academic success can be applied to mental tests in at least two different ways. One may correlate the mental test scores with scholarship records as has just been suggested, or one may determine the mental test scores of the students who leave college for various reasons in comparison with the mental test scores of those who remain in good standing. These two criteria usually give similar verdicts regarding a mental test, but they do not always coincide. A mental test may have predictive value for retention in college without being closely related to scholarship grades; such as the general technical information test which has been given, together with an intelligence test, to the freshmen in a large number of engineering colleges. The technical information test contains only extra-school items which an interested boy, during the high-school age, obtains by inquiring about the mechanical and electrical things in his environment. It is reasonable to suppose that the students who have acquired considerable general technical information on their own initiative will make better engineering students than those who have not availed themselves of such opportunities. This is verified by the fact that among those who make high scores in this test there are fewer students who leave college than among those who make low ones. However, the test does not correlate satisfactorily with the freshmen scholarship records which are heavily loaded with mathematics and physics and do not call for the general technical information of the test. These results will be reported more in detail elsewhere. My present purpose with the illustration is to show that academic success can be used as a

criterion for determining the predictive value of a mental test either by correlating the mental test scores with scholarship records or by studying the average mental test scores of the students who leave college in comparison with those who remain.

The correlations between scholarship grades and mental test scores vary from zero to +0.60 depending on the test used, the exact form of the criterion, and the particular group of students tested. In my experience with mental tests for college students I have never seen a correlation between test scores and scholarship grades exceeding 0.60. In fact they rarely exceed 0.5, and 0.4 or 0.3 is much more usual. Those who have been accustomed to work with mental tests for children are frequently disappointed at finding that the correlations for college students run considerably lower, but this seems to be universal. The reason is probably due to the fact that the success and rating of a college student depend on many factors besides his ability, whereas with children, intelligence plays a more exclusive rôle in determining school success. An intelligent college student will frequently have relatively low scholarship grades on account of lack of interest, social distractions, scattering of his talents, college athletics, financial and other emotional disturbances. These factors are usually absent with school children in the grades.

A still more fundamental reason for low correlations with college students as compared with those for school children is the fact that the school children represent a wide range in abilities, whereas college freshmen are a relatively select group representing a high, but rather narrow range of abilities as compared with the range of intelligence for the total population. It is well known that the correlation coefficient is reduced by restricting the correlation table to a narrow range of abilities. If intelligence tests were given to a sampling of one thousand individuals selected at random from the entire population and if all of these individuals so selected were attempting to do college work the correlation between their mental test scores and their scholarship grades would run very much higher than it now does. I have seen mental test enthusiasts plot a correlation table for scholarship and test scores and proceed to explain away the cases in the wrong quadrants in the light of their personal knowledge of the students. By eliminating the frequencies in the negative quadrants one can raise the correlation coefficients but such doctoring

of the data is, though well meant, scientifically criminal. The possible administrative use of mental tests with college students rests not only on the correlations between test scores and scholarship but also on the corresponding correlations between high-school scholarship and college scholarship. Since these correlations usually run lower than the mental test correlations one would be justified in combining them in an effort to increase the reliability of the predictions. That will be the subject of a later report.

Six tests were selected from among those which have been tried with college freshmen. These six tests were arranged in the "Psychological Examination for College Freshmen and High School Seniors" which is Test IV in the series of six tests which were devised for engineering freshmen. The following six tests constitute Test IV:

1. *General information.*—These are some sample items from the test:

Slice is a term used in
bowling golf tennis football

The *silo* is used in
fishing farming hunting athletics

The response is to underline one of the four given words. Although the general information test is not a direct test of intelligence, it is an excellent indirect test of that attribute. Other things being equal it is safe to assume that the bright person will acquire unwittingly a greater range of information than the mentally less gifted person. That social opportunities constitute one important factor in the acquiring of general information is, of course, apparent; but it is certain that this factor also influences the score in other apparently more direct intelligence tests. Empirical evidence on this question is difficult to obtain because increased social opportunities and wealth are in general possessed by the mentally more gifted part of the population. An exhaustive study of this question would necessitate the differentiation of that part of the evidence of intelligence which is due to exposure in the environment usually possessed by the mentally superior part of the population. Be this as it may, the fact remains that a general information test does serve to differentiate more or less roughly the students who

are able to succeed in college work from those who are, for various reasons, unable to survive their freshmen career. The differentiation is only a rough one, and so is every other known criterion for student selection.

2. *Analogies.*—This test is given in this examination in a form different from that in which the analogies test is ordinarily given. These are some sample items:

Underline two words with the same relation as *eraser* and *ink.*
lightning storm water dirt clothes

Underline two words with the same relation as *doctor* and *patient.*
nurse lawyer hospital court client

Inspection of these analogies items will show that they are slightly more difficult in this form than in the form ordinarily used. The customary form of the analogies test gives two words to establish the type of analogy and the set for the answer is practically given by stating the first word of the required analogy so that the candidate only supplies the single missing word. In the present form the nature of the analogy is given by the two words and the candidate must select two words from the given five words which have the same relation as the two given words. This is more confusing especially since the three-word distractors are so selected that they have reasonable associations with both the two given words and with the two required words. It is perhaps more suitable for college students than the easier form.

3. *Sentence completion.*—This is the customary Ebbinghaus test, especially in the form in which it has been developed by Trabue. Since this test is so well known a single sample will perhaps suffice.

The poor............is hungry because............has.............. nothing to...............

This test has good diagnostic value although the correlations between its score and scholarship for college students do not even approach the correlations that are obtained with school children for the reasons previously suggested. The correlations for college freshmen are usually in the neighborhood of 0.35. The chief difficulty with this test is the scoring. Several years ago we made the attempt to prepare lists of acceptable words for each blank space in the sentence. This was used by giving credit for inserts which agreed with one of the samples on the official list

and by counting as wrong any other inserts. This was found unsatisfactory and we later prepared a more extended list with good inserts and poor inserts. The good inserts were given more credit than the poor ones. The scoring of the papers with this arrangement was found by the examiners to be a tedious job, although that should not be a permanent obstacle in the use of the test if its diagnostic value is higher than that of other tests which can be scored in a simpler way. Its diagnostic value is, however, not superior to that of the other tests in this examination and for this reason it was eliminated in the 1920 edition. The test for which the norms are here reported is the 1919 edition which included this sentence completion test.

Perhaps the most serious obstacle in the use of an official list of acceptable inserts for the sentence completion test is that a clever student will sometimes hit upon some novel inserts which make an original and grammatically correct sentence. Such candidates are penalized for their originality. To obviate this difficulty in the scoring we adopted the plan of accepting any good sentence which was grammatically correct and which was, in the main, sensible. Thus the sentence "There are ten days in a week" would not be accepted, because it is not sensible, although it is grammatically correct. However, in the sentence "It is very *easy* to become *well* acquainted *with* persons who *are* timid," in which the italicized words are inserts, there might be some difference of opinion among examiners as to whether the sentence is acceptable because one would not ordinarily say that it is easy to become well acquainted with timid persons. In order to avoid this source of ambiguity, we finally tried to score the test by accepting every sentence which was grammatically correct, irrespective of its content. This seems rather extreme but in practice it seems to be more satisfactory than any of the other methods of scoring this test. In that case a sentence like "There are ten days in a week" would be acceptable, but it is never written. Even if it were written by a smart student it would not score against his intelligence but, if against him at all, on the basis of discipline. The last mentioned method of scoring the sentence completion test is particularly helpful in deciding ambiguities arising from differences of opinion among examiners. As a matter of fact it does not alter the correlations seriously to modify the scoring of this test but even with the last men-

tioned scoring method the time consumed is much greater than that required for any of the other tests in the series.

4. *Syllogisms.*—The instructions are to underline the word *true* if the conclusion is true, and *false* if the conclusion is false. These are some samples of the test:

Since all metals are elements, the most rare of all the metals must be the most rare of all the elements.

True False (Underline one)

We must sell our output either to consumers or to retailers. It is not feasible to sell our entire output direct to the consumer. If we sell part of it direct to the consumer the retailers will not buy the remainder. Therefore we must sell our entire output to the retailers.

True False (Underline one)

The syllogism test is one of the good tests but it is not the fundamentally important test of reasoning ability that we might at first suppose because a very small fraction of spoken and written English is phrased in syllogistic form. A course in logic and familiarity with Euler's circles affect the score in this test markedly but since the test is prepared for college freshmen who have not yet studied logic this does not constitute a practical difficulty. The conclusions for some of the syllogisms are indeterminate and it has been suggested to give three choices in the response to this test, namely, "true, false, indeterminate"; but if a conclusion is indeterminate it is also false in that it does not follow from the premises. Since two responses would in that case be acceptable the test was given with only two alternative answers, true and false.

When the content of the syllogism is known to the candidate it seems to be easier than when the terms are unfamiliar even though the syllogistic form remain the same. This is a good problem and should be investigated by preparing three syllogism tests identical as to the form of the syllogisms but differing in that one should be expressed in terms familiar to the candidate, one in unfamiliar terms, and one in terms of symbolic notation such as letters It would be interesting to ascertain the relative diagnostic value of syllogism tests with special reference to these three types of content. I have already investigated the relative predictive value of syllogisms with monotonous and varied content. The criterion used for this comparison was the scholarship grades of college students. It may be that this is a special case

more general principle that the diagnostic value of a mental enhanced by making it varied in form and content.

5. *Quotations.*—The candidate is asked to read a quotation of two or three lines from some well-known author. Below the quotation are four statements, two of which agree, and two of which do not, with the quotation. The candidate is asked to check the two statements which agree with the given quotation. This is an example:

"No great genius was ever without some mixture of madness, nor can anything great or superior to the voice of common mortals be spoken except by the agitated soul." (Aristotle)

Check two of the following statements with the same meaning as the above quotation.

......Genius is essentially hard work and persistence.

......Contented and serene characters are the ones that produce works of genius.

......Genius and insanity have certain elements in common.

......Strokes of genius are likely to come after times of great disturbance or stress for the individual.

Some proverbs are used to advantage in this test; as, for example:

"Long absent, soon forgotten."

Check two of the following statements with the same meaning as the above proverb.

......Far from the eyes, far from the heart.

......Absence makes the heart grow fonder.

......Distance lends enchantment to the view.

......Out of sight, out of mind.

6. *Number completion.*—This is the number completion test which has become well known since it was adopted as one of the tests for the Army Alpha. It was given before the war as a separate test to engineering freshmen with whom it had higher predictive value than with students in other courses. It was retained in Test IV as a representative of the numerical form of problem since the arithmetical problems in the original try-out editions consumed too much time in comparison with the other test items. My original justification for devising the number completion test was that it affords an opportunity to test the candidate's ability to form generalizations. The following is an example of the number completion test as used in Test IV. Write the two numbers that should come next:

60 52 44 36 28 20 ....

The complete instructions for the test are given more in detail in the first part of the examination pamphlet.

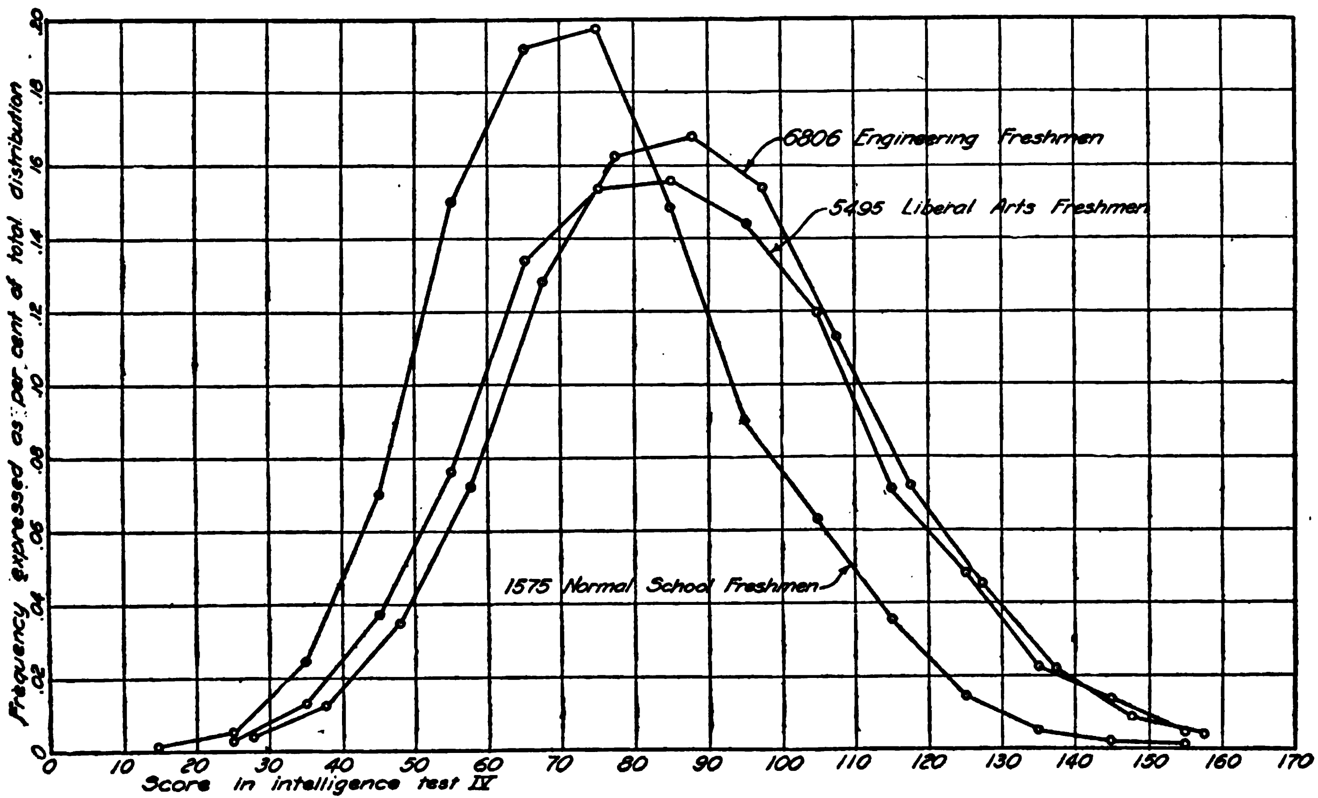


FIGURE 1. DISTRIBUTION OF SCORES IN INTELLIGENCE TEST IV FOR ENGINEERING, LIBERAL ARTS, AND NORMAL-SCHOOL FRESHMEN

TABLE I. DISTRIBUTION OF THURSTONE INTELLIGENCE TEST IV SCORES OF ENGINEERING FRESHMEN IN FORTY-THREE ENGINEERING SCHOOLS, 1919–1920

Schools	Number of Students Who Obtained Indicated Scores																															
	18-22	23-27	28-32	33-37	38-42	43-47	48-52	53-57	58-62	63-67	68-72	73-77	78-82	83-87	88-92	93-97	98-102	103-107	108-112	113-117	118-122	123-127	128-132	133-137	138-142	143-147	148-152	153-157	158-162	163-167	No. of Cases	Av. Score
University of Akron			1		1	3	2	2	2	8	4	5	3	6	8	4	1	4	1	2	2		2	1			1				63	82.6
University of Alabama				1		4	1	4	7	3	5	6	6	3	6	7		5	2	1	1		1	1							64	79.4
University of Arkansas					1		3	2		6	7	3	3	4	4	2	4	2	2	6	1										50	83.5
University of California						1	2	1		3	3	3	9	11	7	12	8	3	4	4	2	1	4	1	3				1		83	95.0
Carnegie Institute of Technology			2	1	5	4	11	20	26	34	36	30	34	50	31	34	28	34	25	15	15	11	8	7	8		2	1	1	1	474	88.2
Case School of Applied Science			1	4		3	5	6	7	6	20	22	22	29	25	23	30	18	18	13	14	17	9	10	6	3	4		2		317	95.6

TABLE II. DISTRIBUTIONS OF SCORES IN THURSTONE INTELLIGENCE TEST IV OF FRESHMEN IN THIRTY-FOUR LIBERAL ARTS COLLEGES

Colleges	Number of Students Who Obtained Indicated Scores																	
	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99	100-109	110-119	120-129	130-139	140-149	150-159	160-169	Av.	No.
Albright College				2	4	5	2	4	5	3	4	2		1				76.6	32
Alfred College					1	1	5	6	13	8	12	4	2	2	1			93.3	55
Center College				1	4	8	9	14	11	3	6		1			1		83.0	58
Coker College		1	3	5	14	25	24	10	10	6			1					61.9	99
Dickinson College				1	3	5	17	19	25	18	15	10	9	5		1		90.0	128
Earlham College			1	1	5	19	21	23	12	29	14	13	10	2	3			86.0	153
Grinnell College			1	1	6	16	21	51	46	40	28	20	11	3	5	2	1	91.1	232
Hanover College	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

College																			
Juniata College				2	1	8	10	9	2	1	7	2						74.0	92
Lafayette College				3	8	21	29	31	43	40	32	20	15	7	2	2		88.5	253
Lehigh University			1	1	3	9	19	24	32	54	45	13	15	5	4			93.0	225
Lincoln University		1	3	4	20	15	13	11	5	4	1	5		1				61.6	88
Muhlenberg College				2	7	8	26	29	29	34	30	11	8	7	5	3		90.5	194
Mount Union College				1	1		5	7	13	10	10	10	3	2	4			97.2	66
Northwestern University			1	6	11	39	76	114	105	111	94	54	31	21	6	5		89.5	674
University of Pennsylvania			1	4	9	13	44	31	38	35	27	33	31	9	8	2		92.0	285
Pennsylvania College				3	4	8	15	14	12	14	8	12	6	3		1		86.3	100
Pennsylvania College For Women					1	6	14	22	18	15	11	7	4	2				86.2	100
University of Pittsburgh					1	5	13	9	7	6	11	6	2	1				85.7	61
Reed College				1	1	3	9	10	16	17	13	16	9	6	4	1		98.3	106
St. Joseph's College				1				2	2	7	9	4	2	2	2		1	107.0	37
Susquehanna University				1	2	4	9	10	10	10	8	5	5	1	2	1		90.0	68
Swarthmore College					1	5	11	20	22	24	20	13	15		3			94.0	134
Temple University		2	1	7	19	37	46	38	32	25	16	11	3	3	2			75.2	243
Thiel College						6	9	15	13	14	14	5	6	6	4	2	1	95.6	94
University of Texas			6	19	56	94	154	161	143	95	65	33	16	8	7	3		78.0	860
Ursinus College				2	5	14	23	28	37	29	25	20	12	2	3			88.5	200
Villanova College					7	10	11	8	17	13	11	2	2	3				82.1	84
Washington University St. Louis				1	7	11	37	67	68	71	85	40	33	19	6	6	1	95.4	452
Westminster College				1	2	8	10	21	16	8	12	2	4		2	1		84.5	87
Wilson College					1	5	17	20	19	19	8	8	9	3				86.5	104
Western College For Women				1	4	8	15	16	18	19	10	5	1		1			82.0	98
Total		4	18	76	209	422	732	846	852	787	657	395	262	123	77	31	4	86.6	5495
Percentile Rank		0	0	1	4	9	19	34	49	64	77	87	93	97	98	99	100		

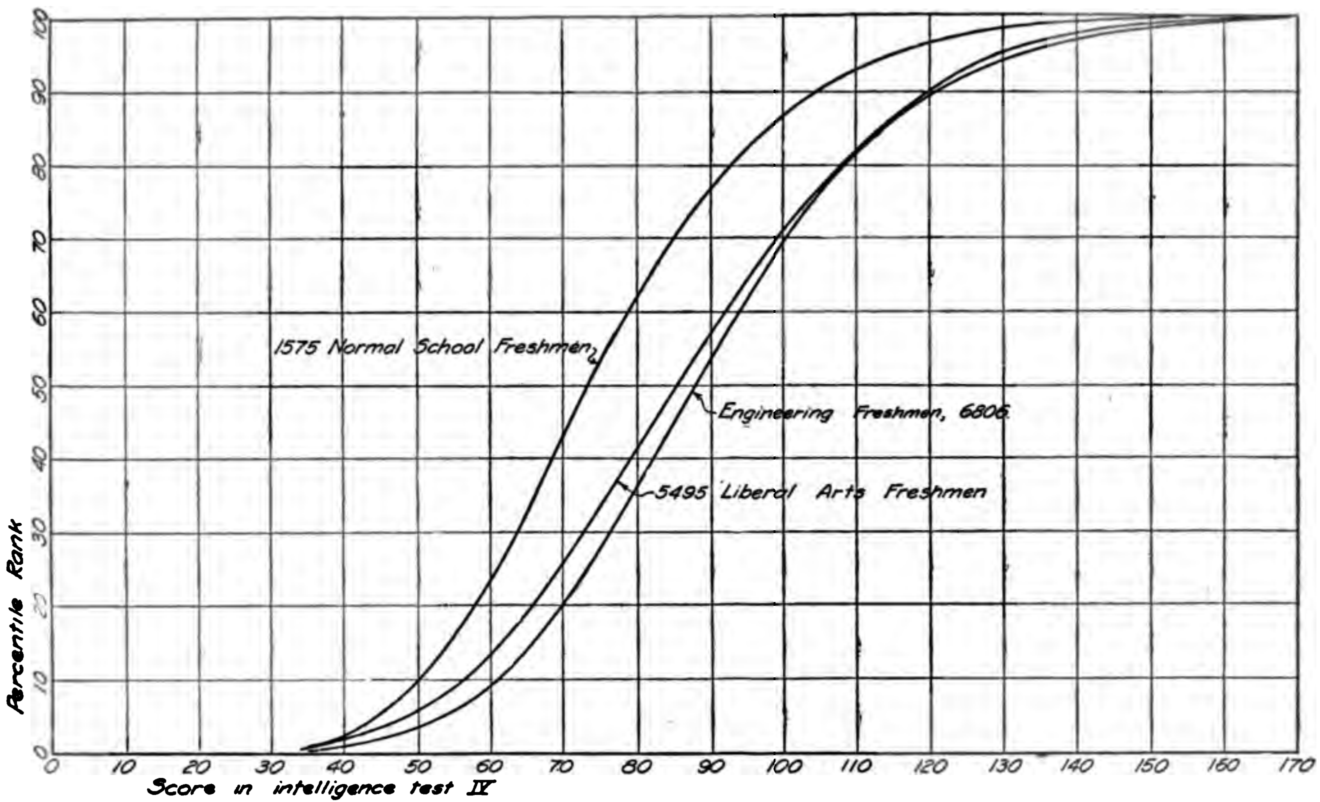


FIGURE 2. PERCENTILE RANKS FOR ENGINEERING, LIBERAL ARTS AND NORMAL-SCHOOL FRESHMEN CORRESPONDING TO THEIR SCORES IN INTELLIGENCE TEST IV

TABLE III. DISTRIBUTION OF SCORES IN THURSTONE INTELLIGENCE TEST IV OF STUDENTS IN TEN NORMAL SCHOOLS

Schools	Number of Students Who Obtained Indicated Scores																		
	0 9	10 19	20 29	30 39	40 49	50 59	60 69	70 79	80 89	90 99	100 109	110 119	120 129	130 139	140 149	150 159	160 169	No.	Av.
Cumberland Valley Normal School			1	8	13	31	64	47	38	23	15	7	4	1				252	74.2
Edinboro Normal School				1	5	7	15	7	12	7	4	2	2	1				63	77.0
Keystone Normal School		1	1	3	16	29	30	30	21	19	6	4	2	1				163	72.1
Lock Haven Normal School				3	3	8	11	15	9	8	7	3	2	1				70	78.7
Millersville Normal School					2	1	6	8	21	10	15	13	4	1	2	1		84	93.5
Slippery Rock Normal School			2	9	15	36	35	40	25	18	12	3	3					198	71.2
West Chester Normal School				7	12	41	50	55	40	19	17	8	4	1	2			256	73.2
Farmville, Va. Normal School			2	4	19	48	41	45	23	9	6	3						200	67.6
Philadelphia Normal School for Girls.					3	6	8	22	20	14	12	13	3	2				103	88.0
Indiana, Pa. Normal School.			1	5	20	31	43	38	26	15	6	1						186	68.6
Total		1	7	40	108	238	303	307	235	142	100	57	24	8	4	1		1575	74.8
Percentile Rank.		0	0	2	6	17	34	54	70	82	90	96	98	99	99	100			

After selecting six tests as the most serviceable for the prediction of scholarship grades of college freshmen our next problem was to determine the manner in which these six tests were to be presented. The customary procedure had been to give each test separately, distributing and collecting the papers for each of the six separate tests. This consumes considerable time and it also complicates the handling of the scores for administrative purposes. For purposes of research the test scores should, of course, be analyzed separately; but when this had been done our task was to combine them in one examination with a single intelligence rating. We were not sufficiently enthusiastic about the method of partial correlation to evaluate the fifty separate tests by this method. Each of the six tests gives correlation coefficients with scholarship grades in the vicinity of 0.30.

If the six tests were arranged in succession in the test pamphlet with one time limit for the whole examination the slower candidates would not give any returns on the last test. In order to avoid this difficulty the six tests were arranged in cycle form. Every sixth test question is, therefore, of the same type. No matter how fast or slow the candidate is, he will give returns on practically the same number of questions from each of the six forms. The test in its final form is, therefore, an omnibus test in cycle form and has been described as a cycle-omnibus test. This type of test should not be used when the diagnostic value of each part is being investigated. For administrative purposes, however, it is far superior to the separate giving of the six tests.

A test of this kind with the items arranged in increasing order of difficulty would be more properly called a spiral-omnibus test. This has been done with the Army Alpha questions by the Bureau of Personnel Research at Carnegie Institute of Technology.

7. *Directions.*—There are two forms in which the instructions for a cycle-omnibus test can be arranged. One may give the instructions in complete form the first time a certain type of test question occurs in the test. The instructions may be gradually abbreviated for the successive repetitions of the same type of question. Thus the number completion test is given with complete instructions the first time it occurs and with the sample answers printed. The next time it occurs the instructions are abbreviated and no further answers are given. Another arrangement of the instructions is to give them sample questions as the

first part of the test which is followed by the test proper without any instructions whatever. Test IV was arranged in the first-mentioned manner. This necessitates a certain minimum amount of instruction material throughout the test but this has been reduced so that the reading time for the repeated instructions requires but an insignificant fraction of the total time of the test.

Tables I-III and Figures 1 and 2 give the norms of performance in Test IV for a number of engineering colleges, liberal arts colleges, and normal schools. It will be noticed that the engineering freshmen and the liberal arts freshmen show practically the same distribution according to the intelligence test with a slight noticeable advantage in favor of the engineering students. The normal-school students do not make as high scores on the intelligence test as the college students. This is perhaps not unreasonable in view of the fact that the standards of selection of college students are in general higher than those for normal schools.

In subsequent reports I shall present in detail the predictive value of the test with special reference to scholarship and withdrawal of students as expressed by the various correlation coefficients.

THE MEASUREMENT OF HIGH-SCHOOL ENGLISH

Edward William Dolch, Jr.
University of Illinois

High-school teachers of English are found almost everywhere to be hostile to any suggestion of scientific measurement of the results of their work. To some the word *measurement* is a veritable red flag.[1] To others, the use of scales is "interesting," but necessarily futile and misleading.[2] And the more that measurement of English work is urged and argued for, the more determined and better organized becomes the opposition.

On the other hand, leaders in education are demanding "justification for the emphasis on English."[3] Their condemnation of the work of the English teachers is based, for example, upon studies which show many high-school seniors with composition ability no better than is usually possessed by grade children,[4] or which discover 93 percent of the seniors in the high school of an excellent system matched in writing ability by an equal number of pupils in the freshman class of the same school.[5] If this, they say, is all that can be attained by three or four years' instruction, let English give way to some subject that can show results. Not that they are hostile to the subject; on the contrary, they are in complete sympathy with what the English department professes to teach. But they are expecting from English just what they are expecting—and finding—in the case of every other established school subject, some definite product of appreciable amount to which one can point and which everyone must recognize. Such a product, scientific study fails to discover in the case of English to an extent that would at all justify the enormous expenditure of pupil time and energy on the subject.

Of course it has been the fashion to say that this disagreement is due to the well-known ignorance of English teachers of the

[1] Ward, C. H. "The scale illusion," *English Journal*, 6:221–30, April, 1917.

[2] Parker, Flora E. "The measurement of composition in English classes," *English Journal*, 8:203–8, April, 1919.

[3] Judd, C. H. *Psychology of high-school subjects*. Boston: Ginn and Company, 1915. p. 134.

[4] *Report of the high-school visitor, University of Illinois, 1919-1920*. Urbana, Illinois: University of Illinois, 1920. p. 46.

[5] Courtis, S. A. *The Gary public schools: measurement of classroom products*. New York: General Education Board, 1919. p. 224.

science of education. Since measurement of educational products is something they know nothing about, their hostility is only natural, and will disappear when they have learned more about the subject. Such a statement is, however, only a half

CHART OF HIGH-SCHOOL ENGLISH

- High-school English includes
  - Literature
    - appreciation
      - of the artist's
        - feeling
        - imagination
        - thought
      - of special technics
        - verse
        - stories
        - novels
        - essays, etc.
    - knowledge about
      - authors
      - books
      - human nature, social history, etc.
  - Inculcation of ethical ideals
  - Writing
    - neatness
      - handwriting
      - papers, etc.
    - conveying of fact
      - conventions of language
      - efficiency in expression
      - technics of special forms
    - conveying of feeling
      - meaning of art
      - artistic use of language
      - technics of literary forms
  - Training of mental processes
    - keen observation
    - logical thinking
    - imagination
    - spontaneity and expressiveness
    - socialization of attitude
    - tolerance, *etc.*
  - Reading
    - greater speed
    - greater comprehension
      - scrutiny of words
      - analysis of thought
    - wider reading habit
      - classics
      - modern novels
      - magazines
      - newspapers
  - Incidental information concerning
    - vocations
    - politics
    - social life
    - science
    - history
    - economics
    - *etc.*, *etc.*
    - from assigned reading, theme subjects, oral discussion, *etc.*
  - Speech
    - manner
      - poise and self-confidence
      - posture, gesture, *etc.*
      - pronunciation, vocalization, expression, *etc.*
    - material
      - correctness according to conventions of language
      - organization of material

truth. The situation is complicated also by a profound misunderstanding on the part of educational investigators. The English problem has been regarded by them very superficially indeed. Experienced in other fields, they have been over hasty to diagnose the situation in English. Inconclusiveness of results and an outcry from the English teachers have been the natural and inevitable consequences. What is now needed is a new attack on the problem from a ground of mutual understanding, and to help establish such an understanding is the purpose of this paper.

There is a definite reason why no scientific study of high-school English yet made has discovered results commensurate with the pupil time and effort given to the subject. Failure of all such efforts is at once understood from a study of the above chart, which attempts to list, in a grouping as logical as possible, the things which the English teacher is supposed to do.

This chart is the result of an attempt to list all the things which the high-school teacher of English is supposed to accomplish in his forty or fifty minutes a day for five days of the week. Very possibly there are other items, but surely these are sufficient to indicate the source of the difficulty. The Latin teacher is teaching Latin, the algebra teacher is teaching algebra, the physics teacher is teaching physics, but just what is the English teacher teaching? Well, his work falls under the fifty-odd headings of the chart. It includes a dozen or more of very complicated skills, several dozen branches of knowledge, and all of those intangible but highly important "attitudes of mind" which make an individual's life a success or a failure. And it is objected that he does not produce *measurable* results.

But to show just how the situation exists in concrete terms, let us consider a typical high-school English course of study. The following is the plan of the course used in the high-school of a large city, and it is typical of the sort of thing done in most well-organized schools. We shall give only the plan for the second year, as those of the other years are very similar.

COURSE OF STUDY IN HIGH-SCHOOL ENGLISH

Second Year

First Half	*Second Half*
I. LANGUAGE	
A. Paragraphs	A. Paragraphs
B. Sentences	B. Sentences
C. Words	C. Words
D. Prosody and figures of speech	D. Prosody and figures of speech

II. LITERATURE	
A. Eliot—*Silas Marner*	A. Shakespeare—*Merchant of Venice*
B. Arnold—*Sohrab & Rustum*	B. Tennyson—*Idylls of the King*
C. Lowell—*Sir Launfal*	C. Shakespeare—*As You Like It*
D. Coleridge—*Ancient Mariner*	D. Dickens—*Tale of Two Cities*
E. Burns—*Cotter's Saturday Night*	E. Cody—*Great Poets*
	1. American
F. Shakespeare—*Twelfth Night*	
G. Hawthorne—*House of Seven Gables*	
H. Cody—*Great Poets*	
1. American	
III. ORAL EXPRESSION	
A. Dissertations	A. Biography of American authors and statesmen
1. American poetry	
2. American novels	B. Periods in American literature
B. Debates	C. Debates
IV. WRITTEN EXPRESSION	
A. Narration	A. Description
V. ELOCUTION	
A. Poise	A. Poise
B. Voice	B. Voice
C. Articulation	C. Articulation

Comparing this course of study with our chart, you will notice at once that certain of the chart headings are included and that others do not at first seem to be. For instance, *Elocution* in the course of study has clearly to do with the *Manner of Speech* of the chart. *Language* concerns the "Conventions of Language" as used in both *Writing* and *Speech*. *Oral Expression* has to do with "Technics of Special Forms," which appears in the chart after the division *Writing*, and also with "Organization of Material," which is placed as a broad heading under *Speech*. *Written Expression*, is, according to the course of study, concerned merely with the technics of narration and description, but it is easily seen that all the habits, skills, and knowledge indicated in the chart after *Writing* must be included in any teaching of narration and description. And this is not all; when you have a pupil write a story, you teach technic of story writing; but you must also in your conferring over his plan and in your criticism of his execution, put time and effort on all the things listed in the chart under *Training of Mental Processes*, and on many more as well. You must stimulate his imagination, guide him in logically thinking out his plot, aid him in observation of people and things, encourage spontaneity, and so on.

But the *Training of Mental Processes* and *Inculcation of Ethical Ideals* must be given the greatest amount of time and energy in the work labeled in every course of study simply as "Literature." Look at the list of classics in this course of study, a list that is quite representative of this sort of work throughout the country, and picture just what is done by teacher and class when these works are being "studied." There is always, of course, some information about the author and what he has written. This practically always includes a sketch of the historical epoch in which he lived as well as an analysis of his life and thought. When volumes have been written upon every one of the writers whose work is taken up, wealth of material is certainly not lacking here, and there is every temptation for the teacher to emphasize vitally important lessons to be learned from the lives of the authors—lessons, however, whose "results" in the pupils' lives are naturally impossible clearly to discern.

Then there is what the chart describes, under *Literature*, as "Knowledge about Human Nature, Social History, *etc.*" What this means for class time and attention in English is best described by the following quotation from a manual on the teaching of high-school literature:

> The class time may be used for the agreeable task of giving the pupils a background for the story. For instance, if the novel is *Henry Esmond*, there is an endless amount of material which can be used to arouse an interest in the picturesque age in which the scene is laid. Queen Anne, Dick Steele, the Pretender, and other characters who flourished in the early seventeen-hundreds; coffee-houses and theatres; brocades, laces, masks, and beauty patches; velvet coats, periwigs, swords and shoe buckles; carved furniture, sedan chairs and stage coaches; all these persons and things afford unlimited means for individual talks by pupils, or discourses by the teacher. The beautiful pictures available for this period will add variety to the study; and many fragments of literature, either old or modern, can be used to supplement the work. The teacher might even read aloud some chapters of that delightful story, *Monsieur Beaucaire*, or parts of *The Bath Comedy*. He might also take pains to show what was going on in America, during Queen Anne's reign, and make some attempt to interest the pupils in the *Virginians*, as a possibility for outside reading.

Pretty hard to show measurable results there, one would think. The educational investigator will at once object that all of this is not English, and of course it is not. It is, as the writer states, "background for the story." It is history, not properly "English." But what is to be done about it? So long as English

must take up the study of books which were written in another age, concerning institutions and events of another age and even another country, just so long must a very large part of English time be given to history. If the scientific student of education wishes to get results in English that are comparable with those in algebra or history, this difficulty must first be solved either by the testing of the history teaching as history teaching, by the correlation of history and English so as to remove this work from the English class, or by the selection of books that will not require that a historical background be taught.

And what is done by the English class with the rest of the literature time? The answer may best be indicated by some questions from a teacher's manual for the study of English classics which is widely used, and the method of which is followed generally whether this manual is employed or not. Under each of the headings only a few questions are here given, these being typical of the whole list. We shall take the questions relating to the first book of those listed on our course of study, *Silas Marner*.

Development of the Plot

How much of the story takes place before the story opens?
How would modern means of communication have foiled Godfrey in his desire to keep some of his secrets?
Why is Mrs. Winthrop introduced?
Where is the climax in interest reached?

Characters

How many principal characters are there?
Are the characters shown most by action, conversation, or the author's comment?
What do you think of George Eliot's power of characterization?

Method and Style

Which is more important in *Silas Marner*, plot or character?
How did George Eliot's scholarship affect her style?
Do the people speak naturally?

It is in such analysis of the books studied that the time of the English class is largely taken up. The various kinds of "Appreciation" listed on the chart under *Literature* are certainly taught; and there is constant *Training of Mental Processes*, especially training in logical thinking. There is, in fact, every opportunity given and used for the development of that capacity which we call general intelligence. But is this English? It is, as English is now constituted; and it will continue to be English unless the senti-

ment of the English teaching force, of the administrative officers, and of the general interested public makes some radical change. But how about tangible, measurable results? There, of course, is the difficulty. At least three-fifths of the English time, and often much more, is devoted to Literature. To trace the results of this time, when the work is of the character described, is indeed a hard matter.

One element in literature study, the *Inculcation of Ethical Ideals*, is so important that perhaps the best-known writer on the teaching of English regards it as the foremost aim of the English course. In this he is followed by many teachers, and there are no English teachers that do not make moral instruction an essential part of all literature study. No matter what the classic, class time or theme-writing time is devoted to discussion of the noble qualities of the characters, to consideration of the ways in which evil character and action deserve and receive condemnation or failure, and to the treatment of those questions which tend to elevate and strengthen the character of adolescent boys and girls. Whether this is properly "English" or not, the study of life as presented by our great authors furnishes an opportunity for ethical teaching that is too good to be lost and one which the public schools rightly feel must be vigorously used. The English teacher of high moral nature and with the ability to impress strongly upon his pupils the moral lessons of literature is felt to be a force whose work is so great as to transcend any methods we have for its measurement. Before this phase of the English course, the scientific investigator stands helpless, yet in any estimate of the whole of high-school English, *The Inculcation of Ethical Ideals* must have full and sufficient consideration.

One heading on the chart, *Incidental Information*, needs a word of emphasis. In English, as in no other subject, there is personal contact between pupil and teacher, and there is the most spontaneous sort of class discussion. There is therefore every opportunity for the exchange of experience and the statement of opinion both by pupil and teacher. In consequence, there is constant comment on current events; there is constant narration of anecdote; there is constant asking of question and giving of answer upon every subject under the sun. The teacher of broad culture and experience constantly emanates information and influence in a way hardly possible in any other classroom. For the content of

English is the whole of life. The themes written concern all of the knowledge, thought or feeling of the young authors; the criticisms of them reflect all the experience and philosophy of the teacher. And the books read, either as classics for the study or from the broader fields of "outside reading," concern all kinds and conditions of men, of all countries, stations, and ages. The English teacher, the pupil, the administrative officer, and the community regard this aspect of English work as of inestimable value. But there is no score card for it. Its measurement is, and will long remain, an unsolved problem.

Is high-school English, then, to be discredited by the movement for the scientific measurement of results in education? Is the scientific investigator yet in a position to say that results do not justify the devotion of so large an amount of time and energy to English work? Hardly so. In fact, the measurement of results in English has scarcely begun. The complexity of the question has not even been realized, and in consequence much trouble has been aroused by premature judgment. There is at present among high-school English teachers a definite and determined hostility to all measurement, simply because educational investigators have rushed to conclusions. This condition is very deplorable, for no great progress is made in any subject except by the cooperation of all parties concerned, and this cooperation is especially essential in a subject such as English. There is enough prejudice about the problems of the English teacher anyway, for the complexity of the subject has already produced numberless violent differences of opinion among teachers. It is unfortunate that the scientist should add still further to the disagreement. He is the one who should have the all-inclusive, the unbiased view. And it is to such a view that this article has endeavored to contribute.

After the English problem is completely understood, then real plans can be made for measuring results. It seems likely that such plans will be based upon some such analysis as is attempted by the chart here shown. The chief point to hold in mind, however, is that analysis of conditions must come first; scientific measurement afterwards.

ANALYSIS OF READING ABILITY[1]

S. A. COURTIS
Director of Instruction, Teacher Training and Research, Detroit Public Schools

A study of the society's yearbook[2] will convince the most skeptical that reading is a complex of abilities, not a unit ability. For some time the outstanding fact in the measurement of reading has been the lack of correspondence between the scores of individuals in a series of tests, labelled reading tests, but very evidently calling for different types of reading responses. Particularly significant have been the disagreements between correlations of scores in intelligence tests and scores in various types of reading tests. Some of these correlations are as high as those between intelligence tests themselves, while others are very low. Disagreements of this sort are indications of unsolved problems; but to one who is directing constructive studies in developing better methods of teaching reading, some fundamental analysis of the situation is essential as a working hypothesis, even though the analysis be tentative and incomplete. My purpose here is to present the conclusions upon which the Detroit constructive work in silent reading is at present based.

Reading, however, ought not to be considered in isolation. In order to attain the maximum of success in a single subject, there must be unity of planning and coordination of effort from subject to subject. Reading must be defined in terms which will make clear the essential nature of what is taking place in the whole educational process.

The definition I consider fundamental for reading and for all other forms of educational activity, is that the basic function of education is to increase the child's control of his own behavior. This definition affords a basis for measuring the relative efficiency of different methods of teaching. In other words, that experience is most educative which produces the largest growth in control of

[1] Address delivered at the meeting of the National Society for the Study of Education at Atlantic City, N. J., February 26, 1921.

[2] *Twentieth Yearbook of the National Society for the Study of Education. Part II: Report of the Society's Committee on Silent Reading.* Bloomington, Illinois: Public School Publishing Co., 1921.

behavior. If we examine the child from this point of view, we find that he differs from the man in his degree of development in each of two distinct systems of controls. One of these I shall call control of purposing, the other control of mechanisms. In the past we have been dealing with these two forms of control blindly and unintelligently. I believe the time has come when the contributions of measurement to our knowledge of the effects of teaching make it possible to organize our experiences, systematize our planning, and achieve results more efficiently because efforts are directed more definitely and more intelligently toward the goals to be attained. Illustrations will make clear the distinctions inherent in these two new terms.

A man's purposes not only differ from a child's, but he has more control over his purposing. The man weighs purposes in terms of meanings or results to be brought to pass a long time in the future. A child's purposes are more in terms of immediate achievements. The purposes of both the child and the man are built out of the same inborn tendencies to action, but the life of an intelligent, well-educated man is determined much more largely than that of the child by his conscious choice of purposes and his conscious organization of experience in terms of his purposes. Therefore, one function of school training should be to give a child (1) experience in purposing, (2) guidance in the selection of worthy purposes, and (3) development of intelligent control over his purposing. In the past this phase of education has been largely neglected.

Mere purposing without achievement is futile daydreaming. For efficient citizenship, as well as for the happiness of the individual, there must be the development of essential mechanisms by which purposes are turned into achievement and of efficient control over these mechanisms. Many a man means well but is actually a failure because he lacks either the essential mechanisms or the necessary control over the achievement skills by which these mechanisms are used to transform his daydreams into reality.

From the beginning schoolmen have recognized the value of mechanisms; hence the emphasis on the three R's. Reading, writing, and arithmetic are primarily the mechanisms by which purposes are achieved. But they are more than this. For it is through the use of mechanisms that new purposes arise. In popular think-

ing, reading, writing, and arithmetic are also fields of purposing, but to school men they have often been conceived only as a field for the building up of mechanisms. Hence the confusion which arises when we attempt to define or measure reading.

Sometimes our attention is directed to "getting the thought" or the purposeful aspect of reading and sometimes towards the development of reading habits or of the control of the mechanism by which the thought is obtained. The Thorndike-McCall reading test is so constructed that the thing measured is the ability of the pupil to use reading for a specific purpose. In my reading test on the other hand, the plan of construction adopted purposely reduces this "use" element to a minimum in order to measure the degree of control exercised by the child over the reading mechanism. An intelligent child, with poor habits of eye movement and much inner speech, will make a high score in the McCall test in spite of his poor reading mechanism, but my test will show up the presence of his defect by low scores for rate. On the other hand a child with perfect control over the reading mechanism but little intelligence will make a low score in McCall's test but a high score in mine. Evidently reading ability needs definition from this point of view and I suggest the following terms.

When the individual is dominated by no purpose or set other than that of getting a general impression of the content presented in the matter read, I suggest that we call his performance observational reading. The traveler waiting in a Pullman at a station and idly reading the signs on a nearby billboard, a man settling down to a comfortable reading of the Sunday paper, a girl reading a novel for pleasure, a child taking the Ayres-Burgess test as most children take it are all examples of this type. I have elsewhere defined it as reading to get the essential relations between the essential elements.

It must be said, of course, that in actual life pure observational reading never occurs apart from some degree of purposeful reading, that the reading mechanism is never operated as a mechanism apart from purpose. On the other hand, so great is the emphasis placed upon developing mechanisms and control of mechanisms that we have children who can read aloud fluently material of whose content they have little understanding.

A still more striking case of the operation of the reading mechanism as a mechanism apart from purpose is found in the

experience we have all had of reaching the bottom of the page only to find that while our eyes had been diligently seeing the words, our brain consistently reporting the association and ideas related to these symbols, our conscious mental self had been attending to something else. In other words there was in operation a psychological complex of intricate habits coordinated and integrated into a process we call reading. It is to this integrated entity that I wish to apply the word mechanism. In observational reading, the self sets this mechanism at work, then stands idly by watching the moving picture of ideas which the mechanism brings to conscious attention.

Observational reading is characterized by openness of mind, by a passive attitude toward what is read. In observational reading little attention is paid to precision of understanding. It is reading for general impression and emotional response. The amount and quality of the reading is determined almost wholly by the skill of the reader; that is, by the degree of control the child has over the mechanics of reading. Of two children of equal capacity, maturity, and training, the slower reader will be found to have established less efficient basic mechanical habits.

The second type of reading I suggest should be called selective or analytical reading and I distinguish two forms, (1) scanning or superficial reading, and (2) study or intensive reading. Both are characterized by the use of the reading mechanism for a specific purpose. Attention is paid only to certain elements of the situation, those elements being determined by the reader's purpose. In the language of Thorndike's psychology, certain bonds are made ready to act by the reader's set; all bonds which are not pertinent to the set are made unready. Thus a proof reader becomes sensitive to errors and sees little else, the politician who made a speech last night responds to the one article in his paper which gives an account of the meeting, the teacher of English reading the same article may see only the faulty syntax.

In all reading of this type there is carried on simultaneously an inner critical or selective process into which reading enters as one element only. Consequently any score based upon the total situation may be determined more largely by this new organization of thought processes than by the skill in controlling the reading mechanism.

My thesis is that since there are many different types of reading situations, and it is necessary to restrict the term reading to one of them, the term should be used to connote the simplest form of the reading activity, namely, observational reading. In other words, reading to me means the degree of control possessed by the child over the reading mechanism, or the excellence of the mechanism itself.

Many will immediately ask, "Is not most reading in life for some purpose? Should we not rather restrict the term reading to selective reading and devise a new term to mean skill in the controlling mechanism of the process?"

There is no escaping the fact that most reading in life is more or less selective in character. A man picks up the morning paper and scans it hastily to find an article on an important matter in which he was interested. In doing so he is carrying on a reading process of a very specific type. Finding the article, he begins to read it leisurely, using a different form of the reading process. Coming across an item of utmost importance to him he begins to study it, taking notes, generalizing, summarizing, and using still another type of reading ability. Finding that the time is flying, he deliberately and carefully scans the rest of the article using still a different type of ability. Now, so far as my experience goes—which is not very far, I admit—if the man has faulty habits of eye movements, that defect will handicap all his reading; and if by remedial training we help him to establish better habits or in other words if we improve the mechanism, all his other reading abilities will be benefited. How generally this is true I do not know. I do know it is true in certain cases.

There is another reason why I think the term reading should be restricted to skill in the mechanics of reading. I am reading this paper, you are listening. Listening means carrying on a certain complex, critical process simultaneously with hearing the spoken word. Would you describe the process you are now carrying on as reading? It is almost identical with the process you would carry on if you were reading this paper silently and preparing to pass judgment on its truth or falsity.

My point is that it will help to clarify the classroom situation if we restrict the term reading to exercises designed to develop either the reading mechanism or control over the mechanism, and then proceed to develop other names, other exercises, and

appropriate tests for the various complex situations in which reading skill is used as a means to an end. Scanning, proof reading, summarizing, studying would then convey the idea that while reading was involved in these processes the training to be given should be directed towards gaining control over mental processes other than the reading mechanism.

Consider for example reading and studying. One ability involved is the breaking of the content read into constituent elements. A second ability is that of recalling out of one's past experience those which are pertinent to the different elements to give them a full, rich content. A third ability is the determination of the character and degree of relationships between the elements. A fourth is the manipulation of the elements to achieve a desired end which may be organization, judgment, memorization, or what not.

The child who has little ability to study needs to be trained in purposing first of all. If he still has difficulty, he may need training in the control of the reading mechanism. If this is not the difficulty he may need training in some one of the other forms of mental activity listed. Training in analytical thinking is not called reading when it is associated with hearing, or with sawing and hammering. Why should it be called reading just because the particular tools employed are printed symbols? The degree of excellence in analytical thinking attained by any individual will be a direct function of his general intelligence and not of his skill in the mechanics of reading. That is my explanation of the high correlation which McCall's reading test has with scores in intelligence tests. It measures not reading ability in general, but one form of ability to study. This in no way detracts from its value as a test, but limits its value as a diagnostic instrument in determining whether or not a child had adequate control over the reading mechanism.

The child who makes a high score in my test and a low score in McCall's may need to be given a series of life experiences rather than a series of reading experiences. To call McCall's a reading test is to imply that the child's deficiency can be made up by practice in reading. This is not true. The person who lacks basic experiences can never have those experiences given him by any form of reading. Herein lies the source of much confusion of thought and much waste of effort. My plea is that the term reading be

applied to that which can be developed exclusively by reading, namely, control of the reading habit as indicated by the ability to carry on easily and with pleasure observational reading. Analytical thinking, selection, judgment may be developed by reading for a purpose, but they may also be developed by many other forms of activity.

I wish to close by calling your attention to the fact that the reading exercises reported by Miss Heller and me in the yearbook are an attempt to put this analysis to practical use in the classroom. The basic idea in these exercises is that reading should be taught not as an end in itself, but as a means to an end. We have under way a series of self-help practice exercises by which the child may teach himself to read.[3] I should prefer to call them exercises in purposing. Our preliminary experiments tend to show that they are remarkably effective. Next year I shall hope to present measures of their comparative effectiveness both in giving the child control of the reading mechanism, and in developing control over purposing.

[3] Courtis Standard Practice Tests in Reading. World Book Company, Yonkers, New York.

HISTORY IN POETRY

Sue Hutchison Dodd

History curricula have usually been determined by the judgment of specialists in History. Recently, however, investigations[1] have been made to determine what the teaching content would be if determined upon the basis of the different functions it might serve. The determination of history curricula in this manner involves many difficulties. The present study has to do with those met in determining the specific historical references contained in the 118 English poems required for entrance to the University of Illinois in 1918. The term "specific references" as used in this study means dates, institutions, persons, places, and written productions.

Our first plan was to consider the historical references in each word and phrase, but obviously such a plan would involve a study of the etymology of each word and the connotations which have grown up around it. It would be difficult, for example, to determine what historical flavors and fragments of information are connoted in the phrase "ivy-mantled tower" in Gray's lines "In yonder ivy-mantled tower, the moping owl does to the moon complain."

Important as such a study would be, it was discarded because of practical difficulties; and the plan of collecting only specific references was decided upon.

The classification of the items was developmental, growing as the study progressed; rather than systematic, following an accepted expert classification. The classes finally arrived at were the following: character, event, place, social class, symbol, institution, date, document, people, principle, and established fact.

The first problem faced in the collection of specific references concerned that portion of history included in the thoughts, feelings, resolutions, beliefs, and customs of the individual characters

[1] Bagley, W. C. "The determination of minimum essentials in elementary geography and history" *Fourteenth Yearbook of the National Society for the Study of Education, Part 1;* "Possible defects in the present content of American history as taught in the schools" reported in the *Sixteenth Yearbook, Part I,* by Professor Ernest Horn; "Historical information essential for the intelligent understanding of civic problems" reported in the *Seventeenth Yearbook, Part I,* by Professor B. B. Bassett.

and of the peoples treated in the selections. These necessarily escaped enumeration save as they were outwardly expressed in an event, document, principle, or institution. The lyrics and the drama are rich in feeling, but none of this feeling adheres to the specific references. Furthermore, historical references frequently lose their significance when taken out of their setting in a poem. In addition, literary references are likely to be distorted as historical facts. These problems were not solved.

A second problem was that of scoring. A quotation of fifteen lines from "Alexander's Feast" will facilitate the discussion:

> 'Twas at the royal feast for Persia won
> By Phillip's warlike son;
> Aloft in awful state
> The godlike hero sate
> On his imperial throne;
> His valiant peers were placed around,
> Their brows with myrtle and with roses crown'd,
> (So should desert in arms be crowned),
> The lovely Thais, by his side
> Sate like a blooming eastern bride
> In flower of youth and beauty's pride:—
> Happy, happy, happy pair!
> None but the brave,
> None but the brave,
> None but the brave deserve the fair.

Here it would for instance be impracticable to score all references to *Alexander*. There are eight such references, including the title, as follows: *Alexander; Philip's warlike son; hero; his* (three times); *pair* (distributed as references to Alexander and Thais); and possibly, *brave*.

It is quite evident, however, that in their bearing on the importance of knowing about Alexander such references have one meaning if derived from eight different poems and an entirely different meaning if derived from fifteen lines of one poem. In the former case Alexander would be a character referred to frequently in literature; in the latter, only once, though frequently on that occasion. This line of reasoning led to the decision to score characters, places, dates, institutions, and written productions only once for each selection.

A third problem was connected with the scoring of dates,—for example, the disposal of *Written in Early Spring* and *St.*

Cecelia's Day, 1687. The first was discarded because it was not a specific reference to a date and the second was included because it was specific. So, also, were discarded for the same reason: *December*, *Christmas*, *Martian Kalends*, and *Marathon Day*, in the "Battle of Lake Regillus." Dates appearing in the titles of poems were included while dates showing the year of publication were omitted.

Only twelve dates appeared in the 118 poems. These were *Marathon Day, 490 B. C.*, in "Browning's Pheidippides"; *St. Cecelia's Day, Nov. 22, 1687*, in Browning's "A Song of St. Cecelia's Day"; *Drummossie Day, April 16, 1746*, in Burns' "Lament for Culloden"; *1692* and *May 31, 1692*, in Browning's "Hervé Riel"; *1746* in Collins' "Ode written in MDCCXLVI"; *1802* in Wordsworth's "England and Switzerland, 1802"; *1802* in Wordsworth's "London MDCCCII"; *September 3, 1802*, in Wordsworth's "Upon Westminster Bridge, Sept. 3, 1802"; *1803* in Wordsworth's "Yarrow Unvisited, 1803"; *1803* in Wordsworth's "Composed at Neidpath Castle, 1803"; and *September, 1814*, in Wordsworth's "Yarrow Unvisited."

A fourth problem grew out of the collection and classification of references to institutions. Here institutions were regarded as certain persistent, collective ideas of a people, which find expression in organization of a political, religious, educational, or industrial nature. While institutions tend to be political, religious, educational, or industrial, the differentiation cannot always be clearly made. In the "Lays of Ancient Rome," "The Iliad," and "The Odyssey," each phase of institutional life is clearly tied up with all the other phases. This lack of separation made it extremely difficult to collect references to institutions.

England and *Switzerland* in "England and Switzerland, 1802" are references to political institutions; as are also *King* of Scotland, *Roman Senate*, *Duke* of Venice, references to the machinery of government applied to particular instances. References to religious institutions and practices are such as *Mother Church*, in "The Lady of the Lake," as, *Consulting Taghairm* (Augury of the Hide), *Waving the Fiery Cross*. Here also belong references to the machinery of the church, as *Archbishop* in "Up at a Villa—Down in a City." A reference to an educational institution is *Eton College* in "An Ode on the Distant Prospect of Eton College," and the reference to the *schoolmaster* in "Snowbound." Similarly,

the *ploughman* in Gray's "Elegy" is illustrative of a reference to an industrial institution.

A fifth problem arose in connection with specific references to persons, including references to individuals and peoples. The characters fell naturally into four groups: historical, Biblical, legendary, and mythical, with shadowy lines between each group. Historical characters were considered to be those appearing in "The Century Cyclopedia of Names" and Brewer's "The Reader's Handbook," or those whose actual existence was affirmed in the editorial or textual exposition of the author's writings found in special editions. To be tabulated as an historical character the person must have actually lived and must have a definite record in addition to that made in the poem. References in Wordsworth's "La Belle Dame Sans Merci" were not included because only specific names or designations capable of specific identification were to be listed. For the same reason *doctor*, *sergeant*, and others in "Macbeth" were exluded from consideration. Similar action was taken in the case of common nouns, such as *peers* in "Alexander's Feast." Where different names for the same person were used, they were included if they carried different connotations, as, *Macbeth*, *Thane of Cawder*, and *King of Scotland*. Names of authors of poems were not included unless they appeared within the poems. Personal references by pronouns were not included.

One hundred and twenty-seven different characters appear in the 118 poems. Of these a number occur more than once, *Alexander*, *Caesar*, *Cromwell*, and *Shakespeare* ranking highest in frequency. These characters represent twelve different nationalities. They are distributed as follows:

English, 27.—Bariffe, Sir George Beaumont, Princess Charlotte, George Chapman, Oliver Cromwell, John Dryden, Queen Elizabeth, King Edward, Thomas Ellwood, Arthur Golding, John Hampden, Childe Harold (Lord Byron), Samuel Johnson, Kempenfelt, Milton, Thomas Otway, Henry Hotspur Percy, Lord Queensberry, Radcliffe, Shakespeare, Shelley, Siward, Hugh Standish, Ralph Standish, Thurston de Standish, Wat Tyler, Wordsworth.

American (including Indian), 23. John Alden, Aspinet, Chalkley, Corbitant, George Haskell, Stephen Hopkins, Merdy E. Hussey, Harriet Livermore, Priscilla Mullins, Sanroset, Wm. Sewel, Squanto, Miles Standish, Rose Standish, Tohamahamon, Mrs. Mercy Warren, Richard Warren, George Washington, Mrs. Whittier, Moses Whittier, Elizabeth Whittier, Miss Whittier, Gilbert Winslow.

Scotch, 17.—Jean Armour, Burns, Mary Campbell, William Douglas, Ellen Douglas, James Douglas, King Duncan, James V, Lisley, Macbeth, Lady Macbeth, Macdonald, Malcolm, Queen Mary, Mary Morrison, Walter Scott, Earl of Douglas III.

Roman, 17.—Mark Antony, Brutus, Caesar, Saint Cecilia, Cicero, Cornelia, Hadrian, Horace, St. Jerome, Livy, Cecilia Metella, Pompey, Publius Scipio, Lucius Sulla, Titus, Trojan, Virgil.

Italian, 15.—Vitori Alfieri, Alfonso II, Michael Angelo, Ludovico, Ariosto, Boccaccio, Dandolo, Dante, Doria, Laura, Pope Gregory, Machiavelli, Petrarch, Rienzi, Tasso.

French, 10.—Belle Aurore, Boileau, John Calvin, Charles V, Louis XIV, Saint Francois, Jean Francois, Jean Lennes, Napoleon, Herve Riel, Tourville.

Macedonian, 4.—Alexander, Philip, Timotheus, Thais.

Greek, 5.—Apollonius, Homer, Miltiades, Pheidippides. Plato.

Persian, 3.—Darius, Rustum, Sohrab.

Swiss, 2.—Queen Bertha, Francois Bonivard.

Spanish, 1.—Cortez.

Egyptian, 1.—Cleopatra.

Twenty Biblical characters are named. Without reference to frequency of mention these are:

Abraham, Amun, Ataroth, Baal, Bathsheba, Beelzebub, Boaz, Cain, Christ, David, Eve, God, Goliath, Isaac, John, Mary, Og, Paul, Rebecca, and Satan.

Under Historical peoples are included references to those not now in existence as well as to those now existing. They include references to clan, as *Clan Alpine;* references to line of descent as *House of Beaudesert* and *House of Tullibardine;* references to family group as *Douglas and Graeme;* references to religious groups as *Druids*, *Hebrews-Jews*, *Christian* (and under the last caption may be classified *Franciscans*, *Pilgrim*, *Puritan*, and *Quaker*); references to tribe as *Cherokees*, *Creeks;* references to groups by nickname as *Yankee*, *Highland*, and *Lowland;* references to city-state groups as *Athenian*, *Carthaginian*, *Lydian*, *Spartan*, *Theban*, and *Venetian;* references to races and nationalities as *Aeolian*, *Angles*, *Arabian*, *Austrian*, *Breton*, *Celtic*, *Dacian*, *Dardon*, *English*, *Ephesians*, *Etruscan*, *Etrurian*, *Flemish*, *French*, *Gallic*, *German*, *Gothic*, *Grecian*, *Greeks*, *Hebrides*, *Hebrews-Jews*, *Italians*, *Indian*, *Latian*, *Normans*, *Persians*, *Romans*, *Scottish*, *Spanish*, *Trojans*, and *Turks*. In the collection of references to historical peoples, the principle was established of including proper adjectives appearing in such expressions as *Arabic* letters, and *Lydian* measures.

References to persons in "The Lays of Ancient Rome" were, upon expert advice, not included among historical characters. Legendary characters, human beings about whose actual existence there is no authentic record (*Robin Hood*) and mythological characters, superhuman beings conceived of as possessing divine characteristics (*Apollo*) were not included.

With respect to the collection of references to places, difficulties appeared in the differentiation between merely geographical expressions and those of an historical character. A case in point is that of *Persia* in the first line of "Alexander's Feast." Persia, of course, is a geographical expression of the present day, but in addition it is a historical place in the sense in which the term "place" is used in this study because associated in this poem with important historical events.

Nine references to written productions were found: Bariffe's "Artillery Guide," Goldinge's "Commentaries of Caesar," the Bible, the One Hundredth Psalm, and Proverbs in the "Courtship of Miles Standish"; "Chalkley's Journal," Sewell's "History of the Quakers," and "Arabian Nights" in "Snowbound"; and Chapman's "Homer" in "On First Looking into Chapman's 'Homer.'"

It will be observed that the number of historical references in these poems is relatively small. We have, however, from a broad point of view secured but a part of their historical material. Our methods may be summarized as involving the following program:

1. To score only once each reference of a kind in a given poem.

2. To collect only specific references to date, institution, person, place, and written production.

3. To regard as historical only those references to dates that are specific and definite in point of year.

4. To collect references to political, religious, educational, industrial, and social institutions that were not running in 1900.

5. To regard as historical only those references to persons, meaning both character and people, who have actually lived, and of whom we have a definite record in addition to the poem in which mention is made.

6. To regard as historical only those references to place that are associated with an historical event and character.

7. To collect only references to written productions contained within the poems.

8. To include all historical references to date, institution, person, place, and written production, that are unmistakably implied but not directly mentioned.

9. To include all designations for the same person or people, which carry different connotation.

10. To include references to historical character regardless of treatment by poet.

Obviously, this method of determining historical content involves a great expenditure of time and labor. The questions naturally follow, is this necessary? Of what practical value are the findings? A list of the historical findings in English poetry is of practical value to both the teacher of English and the teacher of History. While a history curriculum can not, of course, be worked out on the basis of historical references in poetry alone nor even on the basis of such references in both the poetry and prose, read in the High School, a study of these references will aid in coordinating more satisfactorily the history curriculum with other things which the student is expected to learn in his High School course.

The use of scientific methods is imperative to avoid the overlapping of curricula in history and allied branches, and in order to determine what historical data should be taught. Clearly, a pupil should be taught the history that he will have most occasion to use, and it is only through the use of scientific methods that we can learn what it is. Similar methods may be employed to advantage in working out the historical references to the prose literature taught in the high school. Through such studies a more effective coordination of history and literature may be accomplished, time saved through elimination of unnecessary duplication, and better history curricula obtained.

INVESTIGATIONS UNDERTAKEN BY THE SOCIETY FOR EXPERIMENTAL PEDAGOGY IN DENMARK[1]

CHRISTIAN HANSEN TYBJERG
Hjortholms Alle 21, Copenhagen, Denmark

No foreign correspondent with whom we have come in contact has expressed a more sincere admiration for the work which is being done in America in educational research than has Professor Tybjerg of Copenhagen, Denmark. He is—or was when we last heard from him—President of the Danish Society for Experimental Pedagogy. For this reason as well as for other reasons, he is especially well qualified to speak on educational activities in Denmark. He has written this article at our request. EDITOR.

I wish to thank you for the opportunity of expressing myself in your distinguished journal concerning the work in experimental pedagogy here. I read the JOURNAL OF EDUCATIONAL RESEARCH with the greatest interest, and it always comes as a welcome guest. I am also grateful for the many excellent ideas which the American workers in experimental education are dispensing so lavishly; and as I recently expressed it at the "School Meeting of the North" in Christiana, I consider America the leading country of the world in this field.

As to the work we are doing, our society, the Society for Experimental Pedagogy, was founded seven years ago; and during the time it has existed it has undertaken a number of investigations, both of a physical and of a psychological nature. I shall mention these investigations briefly.

RETENTION AND REACTION IN RELATION TO MENTALITY

One of the first of these was conducted by Dr. R. H. Pedersen. The problem was to determine the difference between the ability of normally gifted children and children of inferior intelligence in retaining impressions. In the experiments visual and phonetic impressions were used. For the visual impressions figures were chosen which were constructed according to a simple principle that had been indicated by Professor Alfred Lehmann. From a

[1] We desire to express our appreciation to Doctor Harold M. Westergaard of the University of Illinois for translating Professor Tybjerg's article.

straight line, lines extend with even spacing; these lines are either perpendicular to the straight line, or they make an angle with it of 45°. In this manner a great number of figures are formed which are all different but which are, nevertheless, so much alike that they are about equally difficult to retain in the memory. In the experiment, series of one-placed numbers were also used, each series consisting of eight terms.

Results.—The children in the regular school showed decided superiority in retaining series of numbers. Then follow the children in the auxiliary school (children of inferior intelligence) and finally, the children of the Vaerneskole (children with the lowest intelligence), although the latter are a little older. As to retaining of geometric figures, the children of the regular school take first rank; but the children of the Vaerneskole are better than those of the auxiliary school. This result is due, of course, to the fact that the children in the Vaerneskole are trained more in visual representation; in other words, to the fact that the training in the Vaerneskole is to a decided degree based on visualization. The girls are considerably below the boys in retaining geometric figures, and among adults women have also been found to be inferior to men in this respect. The result of some tests of reaction was that the children of inferior intelligence reacted much more slowly than those of good intelligence. In all kinds of mechanical work depending on speed, they also fell behind those of normal intelligence. The girls reacted more quickly than the boys.

The Ideals of Children

This investigation was conducted by Professor Alfred Lehmann. The children investigated numbered 4,600. All classes of the community were well represented. In an investigation of this sort the motives stated should be emphasized in particular.

Results.—Interests in one's personal acquaintances decreases as age increases, and does so more decidedly for boys than for girls. Besides, the curve for the boys is far more regular, a condition which showed itself in almost all the investigations. On the other hand, interest in historical personalities increases with age. Boys more frequently chose the father as their ideal (person to imitate), girls more frequently the mother. This result is most pronounced in the country, where not a single girl chose the father

as her ideal. Women of history are hardly ever chosen by the boys, but frequently by the girls. The girls more frequently choose persons in Bible history as their ideals. Only a few ideals are taken from poetic works. The dreams of boys about great deeds culminate at the age of fourteen years. From that time on the interest in warriors decreases rapidly while the interest in peaceful occupations progresses accordingly. Girls lack almost completely interest in peculiar masculine deeds. The children of the middle school[2] are far superior to the other children as regards historical interest, in particular at the ages of fifteen and sixteen. Girls more frequently emphasize the looks of their ideals than do boys. Physical abilities and definite positions in life are of more importance to the boys than to the girls. Interest in intellectual ability increases with age. Boys and girls progress at about the same rate until the age of twelve years. During the three following years the girls stand still, but they reach the level of the boys at the age of sixteen.

In respect to motivation, consideration for one's self decreases with age, while general kindness—that is, consideration for others —increases rapidly with age in the case of girls and with remarkable regularity. Boys on the other hand, are not much concerned with general kindness. The motive of fame is mentioned so rarely by the girls that we may leave it out of consideration entirely. Courage, however, is mentioned somewhat more frequently but not as frequently as in the case of boys. Thus we see that for boys and girls two distinct instincts predominate—the feminine instinct for protection and the masculine instinct for fight. In his further investigations Professor Lehmann reaches the conclusion that the coeducational school tends to reduce these differences.

Spare Time Reading of Children

This investigation was undertaken by School Superintendent Huno. About four thousand children were investigated. They were to tell what two books they liked best, and they were also to state what other books or periodicals they read.

Results.—No single book or author obtained a very large number of votes, Robinson Crusoe for 6.4 percent. Cooper is one of the

[2] The middle school has four grades extending from about the twelfth to the sixteenth years. It prepares for higher educational careers.

most favored authors. His books satisfy the warlike instincts of the eleven- and twelve-year-old boys. The interest in fairy tales decreases with age. It is the books with warlike subjects which dominate for the boys. Between the ages of nine and thirteen years these books represent one-half of the preferred list. During the ages of eleven and twelve years, the instinct for fight reaches its greatest intensity, but not until the age of fourteen is there a decided decrease. The line of development runs as follows: from the world of fairy tales through the instinctive desire for fight, where interest moves from the primitive blood-and-thunder fight to the historical forms of war. In the last we reach the condition of the young man who has found his place in the civilization of modern times.

The children of the provinces (the part of Denmark outside Copenhagen) lag behind. They retain for a long time an interest in childish books. The children of the middle school soon get through with Cooper. Boys have a decided contempt for girls' books but girls to some extent read books for boys. For example, 12 percent of the girls chose books on warlike subjects. The most pronounced sentiment is tenderness (kindness, sympathy); and this sentiment rather soon becomes identified with interest in innocent love stories. Girls read fewer books than boys, but their list is more varied. They have more interest than boys in poetry and less in non-fiction.

Physical Condition of Children and the Effect of the Summer Vacation

This investigation included about a thousand children. It was conducted by Dr. R. H. Pedersen. The children were weighed every two weeks, when they had bathed, from the beginning of May until the end of December. In addition to the weighing every two weeks, tests were made of their muscular strength by means of a dynamometer constructed by Professor Alfred Lehmann. This consisted of a spring with a handle, the spring being placed on a board. On this board a transverse piece was placed to support the hand. In each test five pulls were made.

Results.—The children of the middle school had the greatest weight. They weighed on the average two kilograms more than the children of the elementary school. The children of the Vaerneskole (children of the lowest intelligence) weighed the least,

but the relation was not the same in all age groups. The youngest children in the Vaerneskole had the same weight as children in the elementary school, but the children of more than 9½ years weighed less. As to height, the children in the middle school were ahead of the others. On the average they were 2.3 centimeters taller than the children of the ground school.

Concerning the benefit of the vacation the following facts were observed. The younger children when at home during vacation increased 0.54 kilograms in weight; when they were in the country three or four weeks, they increased 0.74 kilograms; and when they were in the country five or six weeks, they increased 0.87 kilograms.

The older children at home showed an increase of 0.97 kilograms. If they were in the country three to four weeks, the increase was 1.04 kilograms. A period of five to six weeks in the country produced an increase of 1.01 kilograms. The children who had jobs increased more in weight than those who did not have jobs, perhaps because their physical processes (circulation) took place more quickly, and they had an opportunity to get better food.

Intelligence and Number of Children

The following three investigations were undertaken by the writer. In the school with which I am connected there are in all twelve hundred children. The children are placed in the classes according to ability—that is, they are arranged in the classrooms according to their marks in the school subjects. As a rule the same teacher follows the children from the first grade to the seventh or eighth grade. Thus, the teacher obtains an excellent knowledge of the children. Now, if the children in each class are divided according to ability into first, second, and third sections, and if one investigates how many brothers and sisters the children have, the following results are obtained. In the first section of the elementary school we find 328 families with 1,355 children, the average being 4.13 children per family. In the second section of the same school there are 322 families with 1,476 children, giving an average of 4.45 children per family. In the third section there are 321 families with 1,599 children; the resulting average is 4.98 children per family. In the auxiliary school 173 families were represented. The total number of chil-

dren in this case was 941, and the average was 5.44 children in a family. In the Vaerneskole (for those of lowest intelligence) there were children from 189 families; the total was 833, and the average 4.41 children per family. As seen by the figures there is an increase in the number of children per family from the first section (that of the most intelligent children) to the school for children of inferior intelligence (the auxiliary school). In the Vaerneskole, however, where the children have the lowest mentality, the average number of children per family is again lower. The explanation probably is that many of these families have so far degenerated that they are about to die out.

I have reached the same conclusion—namely, that in general those of low intelligence propagate the most rapidly—by investigating all the children from 500 families with reference to their standing in school. When the results are put together, they stand as follows, each figure representing the average number of children per family:

	Average Number of Children	
	First Investigation	Second Investigation
Families of children of normal intelligence (first and second sections in the above classification).	4.29	4.4
Families of children of inferior intelligence (third section and auxiliary school).	5.14	5.4
Auxiliary school alone.	5.44	....

The conclusion is that people of low intelligence progagate the most rapidly. In many of the very large families children follow regularly with intervals of two years. Those of highest intelligence decrease the number of children in order to be able to maintain the standard of living. The investigation showed that there is great uniformity in the intelligence of children within the same family, although it sometimes happens that a stupid individual is found in an intelligent family and *vice versa*. The families of skilled laborers have more intelligent children than the families of unskilled laborers.

Periodic Changes in the Health of Children

The material was obtained by investigating children who had entered school during six different years. The school had an enrollment of twelve hundred children.

Results.—There is a decrease in the amount of sickness from the first grade (the youngest) to the third (children of nine or ten years). Then the curve of sickness increases until the sixth grade is reached, decreasing from this point through the seventh grade. Girls have a greater number of days of sickness than boys. The children of the auxiliary school (children of low intelligence) have a greater number of days of sickness than those of normal intelligence. The girls of the auxiliary school have the greatest amount of sickness of all the school children; and in general the children of this school reach the minimum of sickliness later than the other children. During the year there is an increase from April to July. During August (the month of vacation) the minimum is reached. The amount of sickness then increases during the fall and reaches its maximum during the winter in February. The small children, particularly the small girls, show greater oscillations in the amount of sickness, that is, they are influenced more by the seasons.

An Investigation of Children under Council Supervision

Children under council supervision are criminal or morally corrupt. From the results obtained by investigating seven hundred children it was found that the critical age—that is, the age during which the greatest number of crimes is committed—is thirteen to fourteen years for the boys, and fifteen to seventeen years for the girls. Ninety-four percent of the crimes of the boys consist of theft alone. In the case of the girls 39 percent of the crimes consist of theft alone, 28 percent of theft and immorality together, and 27 percent of immorality alone.

The distribution of these seven hundred children with respect to intelligence and industry is given by the following percents. Intelligence: excellent 1.20, good 16.54, medium 8.64, ordinary 8.66, fair 14.76, poor 50.20. Industry: very industrious 1.16, industrious 17.05, fairly industrious 7.75, lazy 74.04.

Editorials

PROMOTION RATES

Promotion rates are computed in various ways which render it difficult either to make comparisons between them or to make accurate statements about them. Three methods have been particularly in vogue: first, the method of expressing the proportion which the number of children promoted on the last day of the year (or term) bears to the total membership on that day; second, the proportion which the sum of the incidental promotions throughout the year (term) and the number of promotions made on the last day bears to the membership on that day; and third, the proportion which the sum of the incidental and end-of-year promotions bears to the membership on that day when likewise increased by the number of incidental promotions.

The first of these methods is the one most commonly used. Its chief defects are, (a) that it takes no account of incidental promotions, and (b) that it represents the school as equally responsible for advancing every pupil on register on the last day of the year or term. Manifestly, both these defects are serious.

As to the first one, every modern school is trying to develop flexibility of promotion. It is trying to offer to every child the opportunity for advancement as soon as it becomes apparent that he deserves it. In such schools it is recognized that fitness for the next higher grade may be acquired at any time during the year and that to afford but one opportunity for promotion is to make it impossible to adjust the school to the child in any but the crudest way. A slogan for such schools would be, "Every day is promotion day."

The second defect of the first method is likewise serious—the defect in virtue of which the school is made equally responsible for the advancement of every child who happens to be on register on a certain day. It is clear that no such equality of responsibility exists. The child who is on register on the last day of the year may have been a member of his grade but a single

day; or he may have been a member of it an entire year. Indeed, if he is a repeater, he will have been a member of the grade in question for more than a year.

The second method takes account of the incidental promotions; but, like the first method, it bases the rate of promotion merely on the end-term membership, thus again tacitly assuming that every child is equally deserving of advancement. Moreover, according to the second method the dividend in the computation may be, and often is, greater than the divisor. Under these circumstances we obtain a percent of promotion in excess of one hundred. Thus, this method while avoiding one of the defects of the first method does so by introducing another; and it fails altogether to avoid the second defect.

The third method, which consists of adding the incidental promotions to both the number promoted on the last day and the number on register on that day, is obviously an attempt to prevent the percents as computed by the second method from exceeding one hundred. This, however, is its only virtue—and that virtue is a vice, because it is obtained by unwarranted means. For example, to add (as is done in the divisor) promoted pupils to pupils on register is to produce an impossible statistical hybrid.

The devising of a valid method for computing promotion rates is suggested as an important research problem. No one seems to have appreciated the seriousness of the condition under which we labor. The fact is that the school has no defensible method at its command by which to record its success in adjusting itself to the needs of the pupils. No one knows what the responsibility of the school is in this regard. No clear case has been made out by anybody to show what should be its total maximum service to the community in regard to promotion—a service which should be statistically represented by one hundred percent. There are those who assert that this maximum service is rendered when every child is promoted. But surely this does not mean that a child who has been in attendance but a day or a week or a month prior to the regular cataclysm at the end of the year or term should be catapulted into the next grade. The problem is exceedingly complex; and before we can determine upon a method of computing promotion rates, we must know and be able to express numerically what the maximum promotion rate (one hundred percent) means.

Moreover, there is little appreciation of the importance of the promotion rate as indicating the chance which a child has to be influenced by the curriculums of the higher grades. Indeed, in our thinking of what constitutes a good school we do not habitually regard the promotion rate as an important element. Yet it is certain that few measures would more accurately reflect the service of the school to the community than would a properly computed promotion rate.

Let us consider an artificially simplified situation. Suppose that a thousand pupils entered the first grade last September in a school system in which the prevailing promotion rate was eighty percent. Consider that none of these thousand pupils will withdraw from school this year—i.e., that all will remain until "promotion day" next June. It is clear that if under these circumstances a promotion rate of eighty percent is applied to these children, eight hundred will enter the second grade next year. There they will join some repeaters; but let us center our attention upon the eight hundred. If we assume that no withdrawals take place during the second year, the promotion rate of eighty percent will reduce the 800 to 640 upon entrance to the third grade.

If we similarly consider grade by grade the survivors of the original one thousand children, we shall find that only 168 of them will be graduated from the elementary school on time—i.e., in eight years. This is truly an astonishing result. We do not mean that, with a promotion rate of eighty percent, only 168 out of a thousand entering pupils will be graduated. Some of them will be graduated in nine years and still others in ten; perhaps some will be graduated in less than eight years. Nor do we mean to say that a promotion rate of eighty percent will have precisely this effect even with reference to on-time graduates, because in any given grade there are both repeaters and pupils promoted from the lower grade on the last promotion day, and because the general promotion rate—derived as it is from all these pupils taken together—may not be the actual promotion rate for each of these two types of children. Our statement is that if one thousand pupils enter the first grade, and if to them and their successive survivors (no promoted pupils dropping out) an eighty percent promotion rate is applied, only 168 will be graduated in eight years.

If, however, a ninety percent rate prevails, the figure corresponding to 168 will be 430. Merely by increasing the promotion rate one-eighth, we may graduate two and a half times as many pupils at the end of eight years. In other words we may more than double the probability that a given child selected at random at the time of beginning school will be graduated on time.

If a rate of ninety-five percent is attained in a school system, 664 of the thousand beginning pupils will be graduated in eight years. This is about four times as many as will have the same chance in a system where a rate of eighty percent prevails.

The reader can verify these figures very easily by taking the proposed rates and applying them eight times beginning with a base of one thousand. One naturally supposes that if the promotion rate is increased from eighty to ninety percent, it only means that ten more children per hundred are promoted. What one fails to apprehend is that this rate is cumulative, that it is not merely applied once but that it makes its deadly assault upon the children eight successive times, and it thus piles up its effects like compound interest—only in the opposite direction.

Is it not worth while to increase the promotion rate just a little in order that such large returns in pupils brought under the influence of the richer curriculums of the higher grades may be realized? Is full service to the community being rendered by the school which—perhaps in the name of maintaining high standards—ruthlessly shuts out from its full benefits the greater part of the pupils who enter its doors?

B. R. B.

THE LATIN INVESTIGATION

Somehow we haven't habitually attributed to teachers of Latin and Greek a decided penchant for new things. We have rather looked upon them as the conservators of our best traditions. Indeed, we have known some of them who were so wedded to the past that their faces did not seem even to be turned toward the light. These, of course, represented the extremists, a fair proportion of whom is to be found in every group. But even the moderate and more liberal of the classical people have been thought of as especially inclined to point out the lessons of the

past. Unusually conclusive evidence has appeared to be required by them before they have been willing to accept new ideas.

The fact, therefore, that a thoroughly modern and remarkably extensive piece of research—with controlled experiments, tests, 'n everything—is being launched by the classicists may mean either of two things, and most probably a little of both. On the one hand, it may indicate that the conventional judgment regarding the attitude of the teacher of the ancient languages toward new ideas is wrong. On the other hand, it may mean that educational research—the modern kind with tests, measurements, and statistics—has proved its case, and that it has done so by the verdict of the conservatives.

Be this as it may, it is a fact that a committee of the American Classical League with the financial backing of the General Education Board is putting on the most progressive, the most fearless, and the most complete program of investigation which has ever been attempted by any group of teachers. The courage of the committee and particularly of the Special Investigators (Doctor Mason D. Gray of Rochester, New York, and Doctor W. L. Carr of Oberlin College) is the more evident when one considers where the money comes from. The publications of the General Education Board have not been characterized by a friendly attitude toward the teaching of the ancient languages. Perhaps this would be denied by the authors of these publications, but it is certain that no classicist would apply to this Board for support in a propaganda to extend the study of Latin and Greek. Under these circumstances one wonders whether this appropriation may not have been made with the belief and possibly with the hope that the results would be in accordance with opinions previously expressed in reports emanating from the General Education Board. In other words, it may have been the intention of the donors to let the classicists hang themselves.

Now, no one is called upon to hang himself merely because he has the opportunity; nor is he called upon even to place himself in jeopardy. The classicists would not therefore have been very vigorously criticized if they had refused to place their necks in proximity to the noose. The fact that they were willing to take whatever risks were involved indicated a high degree of courage. This is, of course, admitting that the attitude of the General

Education Board toward the teaching of Latin is possibly correct and that, accordingly, it may be corroborated by the investigation.

But wouldn't it be interesting if the event should be of a different sort? Suppose, for example, it were found that the study of Latin did, more effectively than any other agency, improve the knowledge and use of English. Suppose it should be scientifically demonstrated that all or the major part of the desirable outcomes claimed for the study of Latin are actually obtained under present conditions or that they are obtainable under conditions which the investigators may set up. And suppose that a report in which such conclusions were reached should be placed in the hands of the General Education Board for publication—for we assume that the Board will reserve to itself the right to publish the report. We wonder whether under these circumstances the officers of the Board would exhibit the same high purpose that now appears to actuate the investigators. In particular, we wonder whether the report would reach the public in the form in which it was written; whether, in other words, the editors would content themselves with a sympathetic treatment of it and limit themselves to allowable editorial functions. The query is not altogether inappropriate, for in connection with the Gary Survey (the report of which was issued by the General Education Board), it has been rather generally believed that the writers were by no means free to express themselves and that this limitation upon their freedom extended not only to form but also to substance.

We offer to our readers on another page a statement concerning this investigation. We urge all research workers and school people not only to be ready to cooperate wherever cooperation is needed, but to be alert to insist that the results of the inquiry shall be impartially reported, that the proponents of Latin shall make no unwarranted claims for it, and that its enemies shall not garble the report in the direction of their desires. We all want to know the extent to which Latin is worthy of our confidence; and we are delighted that a serious attempt is being made to find out. We are grateful not only to the Classical League for accepting the challenge, but to the General Education Board for giving the material support which will make the inquiry possible. And now, just as the investigation is being started, we serve notice upon all and singular the parties thereto that we propose to exercise every

proper prerogative of an interested public to see that the game is played fairly.

We realize the delicate position in which the Special Investigators are placed. On the one hand they are the representatives of the Classical League, either appointed by it or by a committee of it. In everything that they have done or said thus far, they have shown a relentlessly judicial attitude. They are apparently asking no quarter. They appear to be willing to follow the figures wherever they may lead. Whether the Advisory Committee and the influential members of the Classical League are prepared to adopt the same impersonal attitude, especially should the results prove to be unpalatable, is problematical. It is not unthinkable that if the classicists are disappointed in their hopes, they may adopt the tactics of the Turkish sultan whose amiable custom it was to execute the messenger of ill tidings.

On the other hand, the Special Investigators represent the General Education Board with its known proclivities. They will therefore have to prove every point "up to the hilt," especially if the points are favorable to Latin as a high-school subject; and in any event they will have to prepare first-rate reports. For whatever of bias there may be among the officers of the General Education Board on this or any other subject, none will deny the technical excellence of the reports which the Board publishes.

But in the largest and truest sense Messrs. Gray and Carr represent the spirit of inquiry, the genius of research. They are the first teachers of any subject belonging to the secondary or higher curriculum to present a broadly conceived program for the evaluation of the work that as teachers they are doing. If their work as investigators is done, as it undoubtedly will be, in the true research spirit, and if in exemplifying this spirit, they proceed in a workmanlike manner, they will ultimately add something to to the literature of research which under favorable publishing conditions may constitute a classic. The JOURNAL OF EDUCATIONAL RESEARCH wishes the highest and best success to them and to all others throughout the country who are engaged in this hopeful enterprise.

B. R. B.

Reviews and Abstracts

KEITH, J. A. H. and BAGLEY, W. C. *The Nation and the schools.* New York: The Macmillan Company, 1920. 364 pp.

As announced in the subtitle, this book is "a study in the application of the principle of Federal aid to education in the United States." More specifically, it is "a collection of fact and argument designed to show that the Nation is, in a very real sense, an educational unit, that the Federal government should assume a fair proportion of the cost of maintaining schools throughout the country, and that there should be established at Washington an adequate agency through which the educational needs of the Nation *as a Nation* may be made real" (p. 7). In pursuance of this purpose there is presented, first, the historical development in outline of the policy of Federal aid; then comes an analysis of the present situation in the light of the deficiencies revealed by the war, with especial reference to the rural schools and the preparation of teachers; and lastly a discussion, centering on the Smith-Towner bill, of the measures introduced into congress to remedy the situation. The authors present a strong plea in favor of the proposal "to restore the present Federal Bureau of Education to its original status as a Department of the Government, and to make it an executive department with a cabinet officer—a Secretary of Education—at its head" (p. 7).

In the historical part it is shown that, as a matter of fact, the Federal government has always aided education, both by land grants and by grants of money. This policy reaches back more than a century, but at no time has there been any serious thought, on the part of the government, to control public education within the Federal States. The fear that education might become nationalized, in the sense of being dominated by Federal control, has no real basis in fact.

Since the right of the Federal government to support public education has been established by practice, the only remaining consideration is that of expediency. This question likewise has been decided by the course of events. It has become increasingly clear that education is a *National* concern, and that the National interest in education is not adequately cared for under a policy in which the community or the state is the educational unit. The states are independent educationally, as they are commercially and industrially. "If, four years ago, a person could be excusably blind to this essential educational independence of the several states, the time when such blindness is excusable has certainly passed" (p. 320).

Experience has shown that grants of land or money may be used effectively to stimulate the states to greater efforts in meeting educational needs. The money for the purpose can be raised by means of the national income tax—"a method which, in view of the relation of education to the increase and security of wealth, commends itself as eminently fair and right" (p. 322). The creation of a Department of Education is necessary in order to coordinate the various educational activities of the national government, to represent this country in its educational relations with other countries,

to become a national center for educational research, and to furnish leadership in American education. The creation of such a Department would be in line "with what every first-class modern nation except America has already done, and with what the Nation has done in creating Departments of Agriculture, Labor, and Commerce" (p. 323).

Whether or not the reader agrees with the conclusions that are drawn, he is likely to concede that the book is of unusual merit. It is excellently conceived and excellently written. The statements and arguments are backed up with numerous tables and statistics, which, together with the historical *résumé*, are valuable independently of the purpose of the book. The volume is a contribution of note to a fundamental issue in education.

B. H. Bode

Ohio State University

Goddard, H. H. *Human efficiency and levels of intelligence.* Princeton, New Jersey: Princeton University Press, 1920. 129 pp.

That human beings may be classified into a series of widely divergent groups according to the level of their native intelligence; that these levels correspond to stages in the mental growth of children, which are clearly marked according to the various ages; that individuals below the average intellectual level have been arrested in their mental growth at the age which corresponds to their level; that the individual's intellectual level can be accurately measured by means of tests; that the individual can and should be guided, by compulsion if necessary, into an occupation which demands the level of intelligence which he possesses; that crime, being almost entirely due to deficient intelligence, criminals should be dealt with as defectives; and that the education of persons of different levels should vary greatly in range and character—these are the outstanding contentions set forth in the lectures which are here printed.

The author does not attempt in the space of four lectures to give much of the detailed evidence which lies back of his conclusions. He sketches his pictures with broad strokes and only occasionally fills in the details. In general he represents the extreme view regarding the degree of importance of the level of intelligence as a determining factor in human conduct and efficiency, and regarding the accuracy of our present methods of testing. In the opinion of the reviewer the facts necessitate a somewhat more moderate view; for example, concerning the importance of intellectual defect in causing crime, and concerning the definiteness with which, by general intelligence tests, we can determine the vocation which an individual should pursue. Further, the author's opinion that ill-paid laborers would not, because of their low intelligence, enjoy any better living conditions than their wages will buy, and that their dissatisfaction is merely the product of the misguided efforts of more intelligent agitators, needs much more evidence to support it than is forthcoming. The book, however, is a clear presentation of one school of investigators of mental ability.

Frank N. Freeman

University of Chicago

Edwards, A. S. *The fundamental principles of learning and study.* Baltimore: Warwick and York, 1920. 239 pp.

The first four chapters of this text contain an unsystematic and repetitious discussion of habit and habit formation. They represent the author's view that habit is the

chief mental process in reference to education. The next two chapters give some account of perception and reasoning, and of logical fallacies. Following these are more or less conventional chapters on learning, transfer of training and memory. The next chapter is a collection of illustrations of methods of making an appeal to the student. Included in the remainder of the book are chapters on attention, feeling, physical and physiological conditions, and the supervision of study.

The book is very unsystematic in its general plan and in the execution of the various parts. There are a good many citations of experiments, but the original studies which are referred to are not chosen with a due regard for their relative importance nor are the results of experimental work summed up in a careful and discriminating way. Perhaps the most useful part of the book is to be found in the practical devices which the author suggests. Teachers may find suggestions which will repay reading the book, but they will hardly get a clear or complete notion of the "fundamental principles of learning and study" from it.

FRANK N. FREEMAN

University of Chicago

Fifth yearbook, National Association of Secondary School Principals. Menasha, Wisconsin: George Banta Publishing Company, 1921. 69 pp.

The yearbook contains the directory of members of the association and the proceedings of the meeting at Atlantic City February 28 and March 1, 1921. The papers presented at the meeting are given in full, and the discussions are reported briefly. Those interested in secondary education will find material of value in the volume.

The president's address entitled, "The Submerged Tenth," made an appeal for a sympathetic teacher and an adjusted curriculum for that portion of the student group who were mentally unable to do the regular type of work. It is perhaps best characterized by one sentence taken from the discussion of science. "The standards which prevail elsewhere must be discarded and progress must be determined solely by the ability to move on."

"The Scope of Moral Education in Secondary Schools" was treated under the three queries, *why? what?* and *how?* The speaker's answer to *why* is that ethical character has been recognized as one of the seven objectives of secondary education. Eight traits were designated under *what*, and the *how* was answered by the terms, precept, example, and practice. The paper is inspirational though perhaps necessarily somewhat lacking in specific plans for putting the program into operation.

"Social Problems in the High School" presents the general problem well and correctly insists that the social activity of the school shall prepare for active participation in democratic society and not simply furnish opportunity for individual display or clique snobbery.

The round-table discussion of "Biology as a Requirement for Graduation" brought only supporters to the floor, while a like discussion of "How to Encourage a High Standard of Scholarship" found both advocates and opponents of the honor society as a means to this end. The ideas presented by both sides are worthy of careful consideration. The topic, "The High School Principal's Greatest Problem" revealed diversity of views as to what it is. One considers it to be the balancing of administrative and supervisory activities, while another believes it to be the non-holding power of the school. Helpful suggestions may be found on the first problem.

"Some Possibilities Arising from the Use of Intelligence Tests" reads well but offers little either new or suggestive to those who have been reading even a small proportion of the literature which has appeared in this field in the last three years.

"The Growth of Character Through Participation in Extra-Curriculum Activities" outlines some principles and tendencies and offers some concrete suggestions for the attainment of desired ends.

The resolutions of the association indorsed the following: better trained teachers, better equipped buildings, larger executive and clerical force, modification of the traditional requirements for graduation, election of deans of boys and girls in larger schools, adequate financial support for maintenance of a system of efficient public schools, and a national honor society in high schools. The resolutions condemned giving of expensive awards to individual athletes and the existence of secret societies in secondary schools.

Judging from the yearbook, secondary school principals seeking professional growth will do well to ally themselves with the organization and attend the annual conventions.

ERNEST J. ASHBAUGH

Ohio State University

PITTMAN, MARVIN S. *The value of school supervision.* Baltimore: Warwick and York, 1921. 126 pp.

We have come to feel that all things which have to do with rural education have the amplitude of time immemorial, the placidity of a summer sky, the inertia of a home-bred custom, and the vagueness of an unfenced landscape seen from afar. Such a conception has suffered continuous tremors and sharp spasmodic shocks during the last few years as a result of the activity of a group of rural radicals like Pittman, Foote, Bennett, and others. This book by Pittman will administer such a jolt as will make the formless mass and blurred outlines of rural education fairly quiver.

It is an admirably written autobiography of the adventures of a versatile supervisor in charge, for one year, of a group of rural schools. The book tells in detail just what an energetic, well-trained individual of keen intelligence and adaptable personality can do and did do to improve instruction in the rural schools.

Out of this year's experience came at least three original contributions to the technic of rural supervision. One of these contributions is what Pittman calls the *zone system of supervision.* The system provides for an unusually intensive supervision based upon a knowledge of the specific needs of individual pupils secured from the use of standardized tests.

A second contribution is the idea of a professional journal for pupils as well as teachers. By means of a newspaper to the children Pittman was able to take the pupils into his confidence, enlist their support, and maintain their morale on a high level. Through this medium and with the aid of physical and mental measurements of pupils he was able to define objectively for teachers, pupils, and parents educational goals in the mental and physical realms, and show pupils how rapidly they were progressing toward these goals.

A third contribution involved the careful evaluation of the worth of the supervision given. This is the era of testing, and the obligation rests upon each to prove his worth. A few supervisors have accepted this challenge. Most have not. Pittman's

experimental measurements yielded tangible proofs that he had been worth not less than approximately $50,000 to the community. The experimental measurements brought a tremendous indictment against the inefficiency of conventional rural supervision. If a competent supervisor competently trained can double the efficiency of instruction, there is hope for the future of rural supervision but not much argument for the continuance of conventional methods.

There are certain possible weaknesses in Pittman's study which need mentioning. Although any reader of his book will be struck by the extent to which he made use of standard tests, no intelligence tests were used. Recent progress in the joint use of educational and intelligence tests indicate that a judicious use of intelligence tests would have materially improved his supervision.

In the second place, it could be claimed that a still more adequate series of tests would have shown the supervision to be less effective than it appears. It might be urged that Pittman was able to secure marked progress in the formal, measurable abilities at the expense of less easily measured traits. In so far as he was able to measure these more subtle traits there was no evidence that they had suffered as a result of his supervision. There was, in fact, considerable positive evidence that these traits had been markedly benefited. But he was unable completely to silence such critics.

Finally, it is difficult to determine just how much of the success of the zone system is due to the peculiar skill of Pittman or the novelty of the experiment, and how much to the plan itself when operated by any reasonably competent supervisor for a period of years. All these criticisms of his study are voiced in his book and he confesses that they can be countered only partially by the data which he has collected.

Withal this book is unique and distinctly refreshing. It belongs clearly to the new order of things. Pittman has earned the right to a position in rural education above those who have not put their ideas to the test. The sooner careful scientific work and workers receive this recognition the better it will be for education.

Wm. A. McCall

Teachers College, Columbia University

Donovan, John J. and others. *School architecture, principles and practices.* New York: The Macmillan Company, 1921. 724 pp.

A quarter of a century ago a young bricklayer begged Professor Chandler to allow him to become a special student of architecture at the Massachusetts Institute of Technology. In goodness of heart Professor Chandler consented. He saw in the impetuous youth the qualitles of leadership, but it is safe to say he did not expect him to become one of the leaders among the schoolhouse architects of America.

Having completed his school days, John J. Donovan went to California as superintendent for Henry Hornbostel and there took charge of the erection of Oakland's city hall. This work was so faithfully and intelligently performed that on its completion he was selected to design and supervise Oakland's school buildings. Only one who has undertaken a task of this kind can realize the lack of scientific data touching this work; yet Mr. Donovan was expected to give exceptional school buildings to Oakland.

Fortunately he was not of the kind to take responsibility lightly.[1] He went about his task with the enthusiasm and thoroughness that insure success. He took counsel

[1] "John J. Donovan has studied school architecture as though his life depended on it," wrote William C. Bruce, Editor of the American School Board Journal.

with the best architects. Among them was John Galen Howard; and we shortly find him doing some of the best school work in the country with John Galen Howard associated as advisory architect.

At about the time Mr. Donovan was appointed by the National Education Association as associate on its Committee on Standardization of School House Planning, he had decided to prepare his data on school architecture for the use of his brother architects. He knew that they were finding it difficult to procure such information; but the war came and book publishing was put aside for other duties. However, he could not give up his purpose and during his leisure moments he increased the scope of his proposed book until it grew from a mere handbook to a compendium of schoolhouse design.

John J. Donovan's "School Architecture" is a book with a big purpose. It is a carefully planned and well-executed endeavor to bring before the educational and architectural world the requirements of the school, and the ways in which twenty leaders in the teaching profession would meet these requirements.

It would need many pages to do "School Architecture" justice. Not since the admirable work of Dr. Dressler has a book been published that shows so clearly the advance in school architecture. Many of the articles which had been issued subsequently to Doctor Dressler's book were pamphlets, inadequate, and far from satisfactory, showing little knowledge of the facts or of their relation to each other. Mr. Donovan's book shows many of the best-planned larger school buildings. It is a clear and concise presentation of the demands of the large school. In addition to illustrations in the text it contains one hundred and thirty-three pages of plans and photographs showing noteworthy school buildings which have been built during the past ten years. The wealth of photographic and line cuts drawn to scale are of great value because, in most instances, they are drawings of newly executed school buildings which can be visited by the investigator desiring to learn through first-hand observation.

The first chapter of the book deals with sites and grounds and is followed by a chapter on their planning and development. Then come chapters by Mr. E. Morris Cox on the organization and administration of elementary, intermediate, and high schools. These chapters are of great value because they show to the architect and the school man the correlation between the different departments and the ways in which the school organization should govern the plan.

Mr. Clarence D. Kingsley has given exceptionally valuable information regarding the organization and administration of senior high schools as affecting buildings, Mr. J. C. Knight has written on vocational schools, and Mr. W. F. Ewing has treated the administrative department.

Many of the chapters on the divisions of the school building, such as the class rooms, the school library, the assembly hall, the corridors, stairways, and entrances are written by Mr. Donovan; and these chapters and the chapters on the commercial department, the department for home economics, and the cafeteria are profusely illustrated by plans and photographs showing different arrangements of equipment. The author's personal touch tends to enliven these sections of the book and they radiate some of the energy and enthusiasm which have characterized all of his undertakings.

No book published can be called infallible and there are a few things in "School Architecture" that will undoubtedly be questioned. For instance, some of the suggestions are too expensive to be practical and some are mere fads hardly worth permanent

installation; then again, the chapter on "Heating and Ventilating" shows no understanding of open-air rooms nor any appreciation of their uses and advantages. Most of the chapters, however, present their subjects clearly and without prejudice. The intelligence, interest, and zeal of Mr. Donovan and his collaborators are manifest on every page, and the unassuming and graphic way in which the wealth of information is presented makes the reader continue to scan the pages even after the particular data he desired have been found.

"School Architecture" is a book for the investigator who wishes to study what is being accomplished in the details of school planning by the best men of the architectural profession, the men who are working with the educator to correlate the school building with the work of the school child. FRANK IRVING COOPER

Boston, Massachusetts

ROBINSON, EMILY AND JOHNSEN, JULIA E. *Vocational education*. New York: The H. W. Wilson Company, 1921. 359 pp.

This book is a compilation of papers, magazine articles, addresses, and excerpts from books dealing with the various phases of vocational education. The first edition was published by Emily Robinson in 1918. This, the second edition, is a revision by Julia E. Johnsen.

There are seventy-seven quotations arranged in eight divisions with the following headings: phases of vocational education for youth, industrial education, trade schools, commercial education, agricultural education, household arts, vocational guidance, supplemental material for second edition. In addition to the eight divisions of quoted matter there is a very extensive bibliography arranged under the following headings: bibliographies, agricultural education, commercial education, household arts in education, industrial education, trade schools, cooperation of agencies for industrial education, vocational education, vocational guidance, vocational surveys, re-education of the disabled.

The bibliography is very complete and is well arranged. It is the kind of bibliography often needed by students and administrators of vocational education and is the most valuable part of the book.

The use of the heading, phases of vocational education for youth, is misleading as all of the material in the section deals with the philosophy or educational basis of vocational education. The quotations are well worth reading by the young student in this field, but it is unfortunate that none of them bears a date more recent than 1916.

In the section on industrial education the articles quoted are of the same general character as those in the first section, and all of them were written before the enactment of the Smith-Hughes Law.

Through an unusual oversight the heading of the section on trade schools is omitted from the table of contents. The section is an excellent collection of material and covers a very wide range of thought in the field. The usual confusion in terminology is in evidence. This is strikingly shown by the inclusion in this section of a discussion of the Gary system.

The sections on commercial education, agricultural education, and household education are unduly brief, and are limited in the range of ground covered. The same may be said of the section on vocational guidance with the added comment that in view of the newness of the subject and the recent developments in this field, the dates borne by the quotations given make them practically worthless.

Aside from the bibliography, the only part of the book which can be of particular interest to readers who are actively engaged in the field of vocational education is the last section which gives several well-selected articles written since 1917.

It is much to be regretted that in the revision much of the older material was not discarded and each section brought up to date. The past four years of experience, dating from the enactment of the Smith-Hughes Law, have caused a very widespread revision of opinion on the problems of vocational education and any book confining itself mainly to generally accepted ideas of five, eight, and twelve years ago as this volume does, must suffer the criticism that it is somewhat out of date.

A book performing the service this book was intended to perform has long been needed.

A. B. Mays

University of Illinois

Kirkpatrick, Edwin A. *Imagination and its place in education.* Boston: Ginn and Company, 1920. 214 pp.

Recent educational literature has tactfully avoided discussions of the nature, value, and training of imagination. This is due in part to the inability of psychologists to agree on what is and what is not imagination. Again, the mental processes known as perception, memory, problem-solving, habit formation, *et cetera* have so filled the educational psychology of today that imagination has been relegated to a nebulous background. As a result, few parents and teachers have more than a hazy notion of the rôle which imagination plays in the mental development of the child.

This book attempts not only to distinguish imagination from other forms of mental activity, but also to determine something of the part it should play in the educational program. The author discusses "the varieties of imaginative activity in the adjustments of daily life, the changes in the content and form of imagination that occur in the course of a child's development, individual differences in the prominence, intensity, and quality of imagination, and the proper utilization of this activity in the work of the school."

Professor Kirkpatrick's book is the outgrowth of the results of tests given his students in the field of imagination as well as the reports of their introspective studies. No mental activity is more individualistic and less tangible than the imagination, and any attempt to collect data in this field in a scientific fashion should be given due consideration.

The concrete illustrations and examples of imagination which the author presents to support his views serve to make the book more accessible to the lay reader. The language is devoid of technical terminology since the book is intended primarily for use in teachers' reading circles.

The third part, which deals with "School Subjects and the Imagination," seems to be a rehashing of much of the material found in our special-methods texts, clothed in terms to fit the subject at hand. Educators interested in the movement for visual education would do well to read what Professor Kirkpatrick has to say on the value of pictures, moving and still, in training the imagination.

The exercises which occur at the end of most of the chapters are for the most part intended for individual or group study. A few of them point to interesting experiments which a study circle could carry forward. A comprehensive bibliography adds to the usefulness of the volume.

Dean McClusky

University of Illinois

EDMAN, IRWIN. *Human traits and their social significance.* Boston: Houghton Mifflin Company, 1920. 467 pp.

While this is not primarily a book dealing with educational problems, much of its content is identical with subject-matter essential to courses in educational psychology and the philosophy of education. The book was written, the author informs the reader, "originally and primarily, for use in a course entitled 'Introduction to Contemporary Civilization' required of all freshmen in Columbia College." This title might be misleading, for there is no portrayal, as might be expected, of modern institutions and customs, but rather a description of the needs and impulses which give rise to the products and achievements of civilization. The subtitles of the two unequal parts into which the book is divided are somewhat more aptly indicative of the contents.

The first and by far the larger part bears the title, Social Psychology. It comprises eleven admirable chapters dealing chiefly with the essentially and persistently, though greatly modifiable, instinctive nature of man, with special reference to man's social impulses; the birth, significance, and limitations of reflection; the demand for privacy and individuality; the development of the self; individual differences, including sex differences and a discussion of the factors of heredity and environment; a brief history of language and its twofold function of expressing ideas and conveying emotion; and finally, the racial and cultural continuity.

This part presents a masterly survey of the substantial results of the scholarly work of the following authorities: James, McDougall, Thorndike, Dewey, Woodworth, and Trotter. The author avoids controversial matters and technical discussion, being particularly interested to give the student and general reader an easily assimilable, yet thorough, working knowledge "of the fundamentals of human nature and a sense of the possibilities and limits these give to human enterprise." In this the book should be highly successful, for here, as well as in the second part, the treatment savors not in the least of the "textbook" style. It is clear; fresh (even in the discussion of basic instincts); rich with illustrations drawn from the world the student lives in; interspersed with epigrammatic statements, striking and easily retained, which sum up previous discussions; and illumined with a wide variety of well-chosen and rather uncommonly extended quotations which cannot but lead the more reflective reader directly to these sources. The book is written, it can easily be seen, for the student, and even what might occasionally be regarded as redundancy has, one feels, a pedagogical purpose.

The second part, entitled "The Career of Reason," is as well written as the first. The author states that it has been used with success in introductory philosophy courses. The interesting feature here is that the vast fields of religion, art, morals, and science, each receiving an extended chapter, are surveyed from the standpoint of their human value. Reason is, of course, taken in the broadest possible sense as the power by which mankind makes itself at home in the world and wins happiness. The author announces his naturalistic viewpoint, acknowledging chief indebtedness to James, Dewey, and Santayana, whose influence does indeed dominate the discussion.

Religion, art, morals, and science are shown to be partial fulfilments of human needs and longings. "Religion arose as one of the earliest ways by which man attempted to win for himself a secure place in the cosmic order." "While man lives and wonders, hopes and fears, feels the clear beauty, the infinite mystery, and the eternal significance of things, the religious experience will remain, and men will find objects

worthy of their worship." "Science is man's persistent attempt to discover the nature of things, and to exploit that discovery for his own good." "Having in his needful business fortuitously created beautiful objects, man comes to create them intentionally, both for their own sake and for the sheer pleasure of creation." "In the enterprise of Morals, man attempts to discover how to control his own nature in the attainment of happiness." It would be difficult to find many books which present in a more thorough, sympathetic, and fascinating way the human significance of these important phases in the "Career of Reason."

There are minor points, most of them matters of arrangement, which attract critical attention. One wonders, for instance, why, in the excellent chapter on what place science has in the economy of human happiness, there is no mention of the feeling, wide-spread in modern literature, that science has perhaps taken more than it has given in that it has revealed a universe, not friendly and spontaneously providing, but one blind and indifferent to human wants and ideals. In another context, in the chapter on religion, this does find a place, and the author himself says (p. 316): "Nature is thoroughly impersonal, and indeed, were it to be judged by personal human standards, it could with more accuracy be maintained that it is evil than that it is good."

Then again, in the chapter on morals, one misses a reference to what is indeed treated to some extent in the first part of the book, but which cannot be omitted in even a brief treatment of Ethics, since it is probably the fore-most present-day problem in ethical theory, namely, how can group unity be achieved while granting individual freedom?

Very conspicuous, because of the author's usual fairness and carefulness, is his apparent identification of conscience with "instinctive caprice" (p. 451). It is in contradiction to a more tenable statement made on the same page, that "Conscience is thus reduced to habitual emotional reactions produced by the contact of a given individual temperament with a given environment." Yet even this is not the only defensible view to be taken of conscience, as may be seen (p. 433) from a quotation from Dewey and Tufts used by the author himself: "The duty of some exercise of discriminating intelligence as to existing customs, for the sake of improvement and progress, is thus a mark of reflective morality of the régime of conscience as over against custom."

But these are, after all, minor considerations, in a book which in its stimulation and interest should prove an example to writers of students' texts, and which in its scope and sweep and organization should be extremely valuable in giving the student and general reader a broad and unified view of what modern scholarship holds man is and what he may become.

C. Krusé

University of Illinois

Peaks, Archibald G. *Periodic variations in efficiency.* (Educational Psychology Monographs, No. 23.) Baltimore: Warwick and York, 1921. 95 pp.

Variations in muscle strength as recorded by dynamometric tests and in primary memory as shown by the reproduction of a number of digits are investigated in this report. The study was made during the school year 1910–1911 at the Manual Training High School of Washington University, St. Louis, Missouri. The data were secured from ten students tested on each school day and from twenty-two students tested once a week. At each testing each subject was given three trials with the Smedley Dynamom-

eter. On each day the highest score was taken as the record. In connection with each testing the date, the time of day, the general character of the day, and the temperature were recorded.

The facts which seem to be indicated by this investigation are as follows:

1. There are three distinct periods of physical and mental growth during the school year, a period of depression from January to March and two periods of favorable growth from September to December, and from March to June.
2. The mental depression seems to occur later than the physical depression, is not as noticeable, and does not last as long.
3. Mental and physical abilities are favorably affected by sun light, and the stronger its rays, the greater its influence.
4. The lowest and highest temperatures have a depressing effect.
5. No day of the week seems to be especially favorable.
6. Both mental and physical efficiency increase rapidly during the morning with a slight decrease around noon. Mental efficiency reaches maximum about 2 P. M., and physical efficiency a little later.
7. Cloudy days if not too long continued are usually more favorable to both mental and physical efficiency than clear days.

This investigation has been conducted away from the artificial conditions of the psychological laboratory and carried into the actual school environment. The author might have gone one step further and used actual school tasks in place of the artificial reproduction of isolated digits. The number of pupils tested is doubtless too small to give great weight to the conclusions. The investigation, however, shows the possibility of measuring periodicity under controlled conditions in the actual classroom.

P. R. STEVENSON

Ohio State University

News Items and Communications

Why H. S. Graduates go to College

This is the problem which W. P. Morgan, president of the Western Illinois State Teachers College at Macomb, is trying to answer. Last June he sent a comprehensive questionnaire to high-school seniors all over the United States. We can readily see how the answers to this questionnaire, which are now being tabulated, will be useful in educational and vocational guidance.

Self-Analysis at Ottumwa

"When you conduct a recitation," asks Superintendent Blackmar of Ottumwa, Iowa, of his teachers, "do you assign the lesson accurately and definitely; do you always call for a report on special topics; do you have in mind some definite purposes?" These and a number of other questions constitute a suggestive sheet which Superintendent Blackmar's teachers receive. He also lists "Some Marks of a Successful Teacher," and "Some Causes of Failure." We should think that the self-analysis which this material suggests would be exceedingly helpful to the schools of Ottumwa.

Bulletins from West Virginia

Consolidation of Schools: Advantages, Cost, Objections is the title of one of the new bulletins which we have recently received from West Virginia. It was prepared by R. I. Roudebush, State Supervisor of Rural Schools. For its size—only seven pages—it presents more forcibly and more concisely the school consolidation problem than any bulletin which we have lately seen. The "before and after taking consolidation medicine" is well illustrated with clear photographs, tables, and graphic representations. Another bulletin, *A Catechism on Vocational Education in West Virginia under the Smith-Hughes Law* prepared by J. F. Marsh, state director, is well worth studying by all who are interested in Smith-Hughes work. For copies of both these bulletins address George M. Ford, State Superintendent of Free Schools, Charleston, West Virginia.

A Difference Method

Professor L. W. Cole of Boulder, Colorado, is using what he calls "the mental difference" in the examination of paired groups of children. In a recent letter he gives an illustration of the use of this method as applied to a superior and to an inferior division of the first grade. The scores referred to in the accompanying table represent the achievements of the pupils in the Cole-Vincent tests. These tests may be obtained from the Bureau of Educational Measurements and Standards, University of Colorado, Boulder, Colorado. The table illustrating Professor Cole's difference method follows.

	Score	Mental Age	Chronological Age
Superior Division....	31.8	7– 1	6–7.6
Inferior Division....	15.2	5–10	6–2.4
Difference..........	16.6	1– 3	0–5.2

Mental Difference (M. A.—C. A.), 9.8 months.

Value of Handwriting Scales

President S. E. Davis of the State Normal School at Dillon, Montana, raises a question concerning the value of some of our scales and measurements. This question was brought more forcibly to his mind after reading Professor N. A. Harvey's *The Psychology of the Common School Subjects* wherein the author states that after very thorough testing he has been unable to determine that any writing scale has effect in steadying the markings of the graders.

We have certainly seen evidence on the other side of this question; but evidently the matter is not thoroughly settled. We should be glad to have our readers submit evidence on either side. Mr. Davis states that he has used writing scales in handling thousands of papers and that he took for granted that there was a steadying effect. Less extensively he has used the Willing Composition Scale and other composition scales and in their use he also took for granted that the steadying element was present. Mr. Davis believes that the solution of this question might be worth an extended study. "Nothing less," he concludes, "would have any serious value." We agree with him.

Predicting Teaching Success

Mental tests are being more and more widely used. Of that everyone is aware. There is a feeling, however, that in addition to native intelligence there are other qualities having to do with emotional and volitional traits which enter so largely into the success of human efforts that an inventory of these qualities should be provided. An example of this is found at the Edinboro State Normal School where Professor L. H. Van Houten, director of the Extension Department, is about to make a survey of the students both by means of the Thorndike intelligence tests and by means of the Downey Will-Profile Tests. Professor Van Houten adds in this connection, "I am particularly interested in the latter because of some possible value there may be in it for determining the personal qualifications, which make for a successful teacher."

If we can secure an instrument of measurement to be used with students in Teacher-Training institutions—an instrument which will enable us to predict their probable ultimate success as teachers—we are sure that such an instrument will be of a degree of importance which it is impossible to exaggerate. We are likewise sure that general intelligence, while it forms an unquestionably large part in the general equipment required for success in teaching, is by no means all there is to the question. We believe that Professor Van Houten is moving in the right direction.

A Study of Class-Size in Chicago

A study of the relationship between class-size and the efficiency of instruction based on controlled experiments was begun at Chicago last fall by Mr. P. R. Stevenson under the direction of Dr. B. R. Buckingham and with the active cooperation of Mr. Ambrose G. Wight, assistant superintendent in charge of measurement for the Chicago Public Schools. Classes from eight elementary and four high schools were used.

In the elementary schools the relative gains in achievement of the large and small classes were measured by standardized tests. These tests were given near the beginning of the first semester, at the end of the first semester, and near the end of the second semester. To eliminate other causes than that of class-size as affecting the efficiency of instruction the following procedure was used: (1) each class was made large (45 pupils or more) one semester and small (35 pupils or less) the other semester; (2) the same teacher taught the same group of children both semesters (except where pupils were added to or taken away to make the class larger or smaller); (3) all pupils were promoted at the end of the first semester; and (4) intelligence tests were given to all pupils in these classes so that when the number of pupils was increased or decreased, the intelligence level of the class could be kept constant.

High-school classes were selected where a teacher could teach two or more sections of the same subject. Intelligence tests were given to the students, who were then divided into large (30 students or more) and small (25 students or less) sections. These sections were approximately equal in intelligence and variability. By this procedure it was possible to have two or more sections each equal in intelligence, and taught the same subject by the same teacher. At the end of the semester each teacher gave the same examination to her sections. The term grade for each pupil was also secured.

It is planned to publish a report of this study in a bulletin of the Bureau of Educational Research, University of Illinois.

P. R. Stevenson

Ohio State University

Research Concerning Civic Education

The Institute of Educational Research has under consideration a study of civic education. Precedent, however, to any precise study, we are taking up two questions: (a) what have been the contributions to such a study by educators, and (b) what is connoted by the term?

There is abundant material at hand. If we go no further back than the report of the subcommittee of the N. E. A. Committee of Ten and confine ourselves to our own country, a very definite extension is observable which involves not merely the rearrangement in emphasis of studies, but as in the project method of elementary schools, even a new conception of what a curriculum should really mean. The citizen conception has so far outrun the governmental relation that training for citizenship is neither a subject nor an activity so much as a conditioning of the goal of educational effort. Even in secondary schools there are new emphases on social subjects and a new time distribution which may tend to crowd out some more venerable figures of the curricular gallery.

This extension, however, had dimmed certain old boundaries. The scope of civics may be determined by the relations to government, but as so many factors other than political are involved in the creation of new governmental commissions and other bodies, and so many new functions are being initiated, it will be desirable, I think, to determine what the limitations really are as distinguished from those of sociology, economics, etc., not so much in terms of formal definition as in an organization of material.

Aside from subject matter is the other aspect of a curriculum in the broad sense—the rapidly developing forms of societies and organizations involving pupil activities and government, the visits and excursions, the outside organizations, and the cooperation of school authorities therewith—all as they operate consciously for civic training. A tabulation of such activities—place, character, etc. is a good thing and has already been well done for one section. We are considering a more careful analysis of some of them in detail as they illustrate certain possibilities for extension.

One definite piece of research we have undertaken. There is considerable waste effort in teaching the foreign-born, because of failure of civic instruction with reference to those particular phases of past experience that hinder adjustment. Races, nations, and provinces have each of them certain conceptions, attitudes, and habits that do not serve in American life, and these must be considered. Others can be built upon in the program for civic training. What the particular conceptions, habits, ideals, or attitudes are, we shall undertake to discover. Adopting a provisional list of headings: racial origin, racial customs and characteristics, occupational factors as determining civic and social attitudes, political points of view, educational opportunities, relation to other races. We are building up from a selected bibliography a series of descriptions under each head. The descriptions and bibliographies will be submitted to leaders among the foreign-born and to students for comment, revision, and criticism. Possibly the headings will be changed and the bibliographies enlarged.

Albert Shiels

Teachers College,
Columbia University

Cost of Supervision

Assistant Superintendent Charles D. Dawson of Grand Rapids, Michigan, has lately made a study of the cost per pupil for supervision and of the subjects supervised in 39 cities ranging in population from 50,000 to 250,000. The data are for 1920–1921. They were obtained from reports of conditions in September 1921.

In his report, a mimeographed copy of which has lately been received, Superintendent Dawson shows two important tables. The first concerns the cost of supervision and the second indicates the subject in which supervision is provided in the 39 cities.

After commenting in connection with the first of these tables on the fact that the cities differ widely in population and in the number of pupils on register at the end of the school year, Superintendent Dawson says: "A casual observation shows also that the cost per pupil for supervision varies considerably. Wheeling, West Virginia spends $3.69 per pupil while Scranton, Pennsylvania spends but $0.65. Holyoke, Massachusetts spends $3.52 and San Antonio, Texas $0.69. Davenport, Iowa spends $2.89 while Nashville, Tennessee spends $0.70. By noting the school populations of these cities we observe that the three cities, with high per pupil cost, are small; while the cities with low per pupil cost are large. The question arises at once: Does it follow that the larger the school population, the smaller the cost per pupil for supervision? By figuring the correlation (Spearman's rank method) between the number of pupils belonging at the close of school and the cost per pupil for supervision we find this correlation to be −0.77. That is to say, it is true, in the main in this study, that schools having a large number of pupils have low cost per capita for supervision. From these 39 typical schools it would seem that we are justified in concluding that in general the greater the school attendance the lower the cost per pupil for supervision."

The author of this report finds that so far as Grand Rapids is concerned it ranks tenth from the *highest* in number of pupils, but that instead of ranking tenth from the *lowest* (rank 29), Grand Rapids ranks eighteenth. In other words, the cost per pupil for supervision is somewhat high in Grand Rapids.

He points out, however, that this takes no account of the results of supervision and goes on to show the remarkable development of supervision and the satisfactory results obtained from it at Grand Rapids.

In connection with Table II of the report, Superintendent Dawson finds that on city has *special* supervisors for all the subjects listed in the table. The subject most universally supervised by special supervisors is music, all cities but one having a special supervisor of this subject. Also it appears from the table that high-school subjects are least supervised by special supervisors. This is only apparently true, since it is often the custom for superintendents in these cities to reserve the general supervision of high schools for themselves and to assign the special supervision of subjects to heads of departments, who thus become in reality supervisors of special subjects.

In conclusion Superintendent Dawson recognizes the many varying factors which may enter into this question of expenditures for supervision and points out for Grand Rapids that in order to give that city the rank of tenth from the lowest, which a perfect correlation between the pupil population and cost of supervision would imply, there would need to be a cut of from $3,500 to $4,500 in the total amount spent for supervision. In this connection, however, he says:

"In view of the fact, however, that the above amount is so small, and also because the results in subjects supervised are first class, it would seem inadvisable to curtail the work of these departments by cutting the supervisory budget this small amount."

Educational Survey of a Supervisory District in New Hampshire

Mr. S. S. Brooks, whose articles on the use of tests in rural schools are appearing in our Journal, has recently become superintendent of the supervisory union of Winchester, Hinsdale, Swanzey, and Richmond, New Hampshire. Profiting by the experience which he gained while at Silver Lake, he has started a survey of this new district.

"During the past two weeks," Mr. Brooks writes, "I have given the National Intelligence Tests or the Kingsbury Group Intelligence Tests to nearly eight hundred school children. I have given the tests, scored them and recorded the results in form for filing besides doing the regular routine work of the district which consists of three small but wealthy manufacturing towns in the southwestern corner of New Hampshire. There are about fifteen hundred school children in the district and sixty regular teachers including those of the two high schools.

Achievement tests in reading, arithmetic, writing, spelling, language, and geography have been given in all the schools and the results are now being transferred to the graph cards. Results so far indicate that the children will average rather low mentally, somewhere between 80 and 90 I. Q. There is a large foreign element here, mostly Poles and Lithuanians. The children generally are up to grade or above in arithmetic but much below in reading, language, and the content subjects. This tends to confirm the opinion formed as a result of using tests in the other district, namely, that arithmetic is overemphasized in most schools at the expense of the other subjects because it is easy to teach and because the children like it. Much time is wasted in drilling on the fundamental operations after the children are already above standards.

"This preliminary survey will furnish the teachers and myself with a starting point for the year's work and enable us to measure the progress of pupils and the effects of remedial measures to be introduced as a result of conditions found to exist. In all grades intensive silent-reading class drill covering a wide variety of subjects will form the main feature of this year's effort to improve conditions. The reading drill will be supplemented by a definite campaign of vocabulary building."

S. S. Brooks

Superintendent, Supervisory Union,
No. 25, Winchester, N. H.

Incommensurability of Alpha and Beta

For the Army mental testing program there developed two group intelligence tests, Alpha for those who could easily read and write, and Beta for those who could not. About one-quarter of the recruits took Beta. The other three-quarters took Alpha. Each test gave ratings in terms of A, B, C+, C, C−, D, and D−. In practically every company there were men for each test. When the company commander selected his men for various duties on the basis of these ratings, he, of course, assumed that a rating of, say, B in Beta was the same as B in Alpha. When the problem of the

distribution of intelligence ratings according to nativity was studied[1] it was obviously assumed that the letter ratings by Alpha and Beta were commensurate, for apparently these ratings by the two tests were combined in making up the graphic comparisons.

This problem was studied in the fall of 1918 at Camp A. A. Humphrey where the writer was a member of the psychological board. Of the 1,794 enlisted men who took both Alpha and Beta it was found that only 37 percent received the same letter ratings in these two tests. Of 195 men earning A in Alpha, 34 percent got B in Beta, 22 percent C plus, 5 percent C, and only 8 percent C minus. In other words 69 out of a hundred regular Alpha men who earned a rating of A in Alpha would have lost from one to four points, in terms of letter ratings, by taking Beta instead. On the other hand, it was found that of 228 D minus men in Alpha 50 percent got D in Beta, 13 percent C—, 4 percent C, and 1 percent C plus. Therefore 68 percent of those who earned the lowest rating in Alpha would have been from one to four letter grades better off by taking Beta instead. The men scoring high in Alpha would be penalized by substituting Beta, whereas those scoring low in Beta would be rewarded by substituting Alpha. One thing is certain, the two scales are not nearly commensurate.

This discrepancy between the ratings by the two scales presents some very obvious difficulties when problems are studied in the light of the ratings by the Army tests. The writer experienced such difficulties in a study for the War Department of "The intelligence of troops infected with hookworm versus those not infected," being a study of 13,278 cases, which appeared in *Pedagogical Seminary*, October, 1920. In spite of the fact that the statistical experts of the office of the Surgeon General of the Army devised a scheme for reducing Alpha and Beta ratings to a common denominator, this attempt at reduction never proved accurate. The Alpha ratings will always be Alpha ratings, and nothing more.

What an advantage it would have been for practical as well as for scientific purposes if the almost two million intelligence ratings on file in the War Department had been in terms of one continuous scale whose ratings would have had a constant meaning! As a matter of fact a number of men of the Psychological Service were striving in that direction at the time of the signing of the Armistice.

Copying after the Army program, the test makers soon developed group intelligence tests for the public schools, which like Alpha, demanded ability to read and write English rather freely. Hence the earliest of such tests applied only to the upper grades and high school. Soon there came the obvious demand for group tests for the lower grades. Consequently there have developed tests that apply to the lower grades only and other tests for the upper grades only. Moreover, there have been subdivisions, so that there are now tests that apply only to one or two grades.

When the research man wants to use the ratings by intelligence tests to study large school problems involving the intelligence ratings of the children of all grades and ages of a given school system, the same difficulty arises which arises when research is attempted with ratings by the two Army group tests.

Moreover, the same problem which confronted the practical Army man confronts the practical school man. How can he interpret the ratings by one scale in terms of the ratings by another scale? How can he study the ratings of his first, second, third, or fourth grade in comparison with the ratings of his higher grades or high school?

[1] See *Memoirs of the National Academy of Sciences*, 15: 696–98.

From experience with regular school tests and standardized educational measurements, this practical school man has observed that there is a tremendous overlapping, that some children, for example, in grade II do as well in a test as others in grade V. But if the measuring device for grade II does not apply to grade V, how can any one compare the two ratings or compute the amount of overlapping? No one can with accuracy. Of course like the Army statisticians, some test makers have devised schemes for reducing the ratings of their several scales to common units of mental age or to some other units. But no scheme has succeeded in making ratings by any two such scales commensurate and doubtless no such scheme can be found.

GARRY C. MYERS

Cleveland School of Education,
Cleveland, Ohio

The Classical Investigation

A comprehensive survey of the results secured in the teaching of Latin has been undertaken by the American Classical League with the financial support of the General Education Board. The committee intrusted with the conduct of the survey has published a preliminary report which outlines the plan of the investigation.[1] The report is tentative only and the procedure suggested will doubtless be modified in many particulars as the investigation develops.

The problem to which the committee is primarily addressing itself is to determine the content and method by which the aims commonly ascribed to the study of Latin may be most effectively attained and, as the complement of this, to estimate the relative importance to be attached to these aims in constructing a course in Latin for the secondary schools. In forming such an estimate various criteria may be considered such as their relative attainability through Latin, the relative likelihood of their entering into the life-experience of the pupil, and the relative importance of their contribution to the general objectives of secondary education.

By this process it is hoped that a sound factual foundation will have been laid for the discussion and solution of the still more fundamental question as to the place and value of Latin in the curriculum as a whole.

In order to provide an adequate basis for any desirable reorganization of the content and methods of the present Latin course, and in order to make it a more effective and serviceable educational instrument, the committee is planning to inaugurate a comprehensive testing program of sufficient scope both in variety and in geographical distribution to enable it to determine, first, to what extent under present conditions the objectives commonly claimed for Latin are attained and, second, what content and what methods provide the conditions most favorable for their attainment.

The shortness of time available will not permit the committee to await in all cases the results of the general survey before attempting to provide remedial measures. It is a reasonable presumption that the needs for improvement disclosed will be plentiful. Consequently the committee is seeking in connection with every test given in the general survey to institute contemporaneously a controlled experiment in a limited territory in which methods may be tested, results measured, and constructive recommendations proposed. Should the results of the general survey when they are

[1] The report is in three sections. Section A and Section B appeared in the October number of the *Classical Journal*. Section C will appear in the December number.

afterwards obtained prove to be so satisfactory as to render such studies dispensable, the loss will in such cases be cheerfully borne.

Furthermore, wherever the committee is convinced that results could be more adequately secured in certain fields if pertinent data were more carefully analyzed and organized, immediate steps are being taken to provide these data. This applies especially to many of the objectives involving the application of knowledge gained in Latin to other subjects.

In attempting to determine by means of tests whether Latin is performing satisfactorily the various functions it is claimed to perform, the committee was naturally confronted by the absence of any absolute standards of attainment in the various abilities concerned; and, hence, it has been compelled to adopt the comparative method at every stage possible. That is, it proposes to measure the *relative rates of progress* made by pupils who are beginning the study of Latin and by non-Latin pupils of the same grade. It proposes by means of initial equating tests in each special ability tested, supplemented wherever possible by general intelligence tests, to determine what initial superiority, if any, Latin pupils possess and to measure on that basis the relative progress of the two groups.

It is planned to measure these same pupils at intervals during the next two years. The final analysis will, whenever possible, be made upon the basis of those Latin and non-Latin pupils who remain at the end of two years and who can be paired on the basis of their initial scores.

This testing program is ambitious, and can be carried out even partially only through the generous cooperation of departments of research in both universities and municipalities. If the response called forth by the announcement of the first series of tests is a fair criterion, the committee is justified in feeling that a reasonable portion of the program may be carried through. Its plan as presented in the preliminary report is exhaustive for the very purpose of providing as many points of contact and as many fields of common interest and cooperation as possible.

The committee proposes to go wherever the facts lead. No part of the present course is sacred to it. It is a fair statement to say that it is proposed to re-appraise the entire content of the present course in the light of the facts disclosed. The present reading course in Caesar, Cicero, and Vergil may pass the test unscathed, but it will require more than tradition to preserve it. The whole question of the content of the course is to be opened afresh and without prejudice.

The committee is proposing, in other words, to find the facts; and if the investigation results in the exposure of unsightly family skeletons, it does not intend to hesitate on that account. It will be borne in mind, however, that the committee's purpose in exposing any skeletons is to remove them. If the first half of the program is carried out with absolute sincerity, any apparently unsatisfactory disclosures will be but a guarantee that the second half of the program will be prosecuted with equal thoroughness.

MASON D. GRAY, *East High School, Rochester, N. Y.*
W. L. CARR, *Oberlin College, Oberlin, Ohio*
Special Investigators for the American Classical League

National Association of Directors of Educational Research

(E. J. ASHBAUGH, *Secretary and Editor*)

President Rugg recently met with the representatives of the organizations which have been invited by the Department of Superintendence to meet with them at Chicago this winter. He reports that the officers of the department welcomed his suggestion for a three-session program of our association this year. In consequence he has already started to secure the best possible talent for these meetings. In spite of the high standard set in the past, we believe it is safe to predict that those interested in the research programs which will occur on Tuesday, Wednesday, and Thursday afternoons will find that the program has not only been lengthened but also enriched. For the meetings we have been granted the largest room available in any of the hotels. Further announcements will be made from time to time.

Boston.—Dr. Arthur W. Kallom, assistant director of the Department of Educational Investigation and Measurement, reports that the problem of cost in connection with testing has become a real one at Boston. In order that the department may cooperate with the schools to a greater extent the masters (principals) will be required to buy their own tests and test materials. The problem of retardation is being investigated. He says: "The statistics collected in Boston on this matter have not been at all satisfactory, largely because the question of age has been left to the teacher. How far we shall be able to go into the matter, I am not yet certain; but I hope it will be possible to carry the study through to completion." Kallom is also organizing a course in the measurement of intelligence for a group of kindergarten teachers which he believes is going to be both interesting and helpful.

State University of Iowa.—Dr. H. A. Greene, who succeeds Dr. E. J. Ashbaugh in directing Educational Service at the State University of Iowa, outlines the following problems for the coming year. First, they are engaged in a state-wide cooperative survey of arithmetical abilities to serve as a recheck on the survey of arithmetic made in 1915 by Dr. Ashbaugh. The remaining three projects are concerned with silent reading, namely: (1) a state-wide study of silent-reading abilities in the state on a cooperative basis; (2) experimentation with the possibilities of devices for the measurement of comprehension in silent reading; (3) the development of at least two types of silent-reading tests, one for use as a group test in primary grades and the other for use in the upper elementary-school grades.

University of Michigan.—Dr. Clifford Woody, the new director of the Bureau of Mental Tests and Measurements, informs us that 38 public school systems of Michigan are participating in a cooperative testing program launched by the bureau. The

testing programs called for the giving of the following tests during the second and third weeks of October:

National Intelligence Tests, Scales A and B.

Monroe Silent Reading Tests, Form 3, Revised Edition.

Courtis Supervisory Tests in Arithmetic, Forms A and B.

Buckingham's Extension of the Ayres' Spelling Scale (selected words).

In order to connect the achievement tests more closely with the improvement of class instruction, extended questionnaires concerning teaching practices in the various subjects were filled out by the teachers on the days when the tests were given.

When the data for the individual schools have been compiled, they will be forwarded to the Bureau where comparative records will be assembled and state standards determined.

Iowa State Teachers College.—Professor Fred D. Cram has furnished us with an interesting report on a bit of research work he conducted last spring. He calls it "A Tentative Study of the Content of Children's Minds in Respect to Names of People in History." Eighth-grade children were given two minutes in which to write the names of people who have been great in American history. Data were secured from 1,149 children in twenty-three schools. The study was undertaken to ascertain how adequately the subject of American History is being taught and what names are impressed upon the children before they leave the elementary schools.

Two hundred and thirty-two names were mentioned of which thirteen were women. It is interesting to note that Washington and Lincoln were the only names mentioned by more than 50 percent of the children, they being listed by 93 and 87 percent respectively. Seven others, Grant, Jefferson, John Adams, Roosevelt, Robert E. Lee, Columbus, and Jackson were mentioned by more than 25 and less than 50 percent of the children. Of the names of women, two only, Harriet Beecher Stowe and Betsey Ross, were mentioned by more than 10 percent of the children.

The report is exceedingly interesting but is not specific as to the modification of our history work which should be made if this survey reports a truly typical situation

---

West Allis, Wisconsin.—Mr. T. L. Torgerson, director of the Department of Educational Measurements, has furnished us with his Bulletin No. 2 on Tests and Measurements—a bulletin sent by the department to the teachers and persons directly interested in the local situation. The report will be briefed in the next issue of the JOURNAL.

Long Beach, California.—Mr. Ernest P. Branson has favored us with a mimeographed copy of what he calls a Research Primer by the Research Committee of the Long Beach High School. Ten questions are propounded and carefully answered. The questions are as follows: (1) What is intelligence? (2) What is an intelligence test? (3) Of what use is such a test? (4) Of what value is the scheme of having X and Z divisions. (These letters evidently refer to groups of children selected on the basis of intelligence.) (6) What has already been accomplished in the Long Beach High School by the use of the Terman Group Intelligence Tests? (7) Can the same work be given to X and Z sections? (8) Should a Z division enroll as many pupils as

an X division? (9) How should pupils who have been placed in the wrong section be adjusted? (10) Who are members of the Research Committee and what can the faculty contribute to the committee? The idea seems to be a good one for educating the entire teacher group and for informing it as to the activities of the committee.

Grand Rapids, Michigan.—Charles D. Dawson, assistant superintendent in charge of research, has sent in three bulletins. One is a bulletin to the principals, giving report on Monroe's Diagnostic Test on Fundamentals of Arithmetic, given last June. The one thing in the report of perhaps greatest value to our readers is his slogan "Every school up to standard, every grade up to standard, and every pupil up to standard in the fundamentals of arithmetic by 1925." The second is the annual report on the educational status of 35 schools of Grand Rapids as shown by standardized educational results. The report shows splendid progress over the preceding year in each test which was repeated. Mr. Dawson and his co-workers should be congratulated on securing this attainment. The third bulletin deals with a study of the cost per pupil for supervision and of the subject supervised in 39 cities ranging in population from fifty thousand to two hundred and fifty thousand. A brief report of this study will be found under *Communications*.

A REQUEST FOR REPORTS ON CIVIC EDUCATION

The Institute of Educational Research, Division of Field Studies of Teachers College, Columbia University requests from Superintendents, Principals and Teachers:

1. Copies of reports, studies or proceedings of committees of schools or educational associations, local, state, or national, on *CIVIC EDUCATION* or related subject matter.

2. Descriptions and reports on experiments in the organization or administration of schools, or in pupil activities and projects or extra mural activities carried on under school direction or in cooperation therewith.

This request is made in the interest of an analysis of present conditions and tendencies, and of new developments in the field of civic education.

Copies of such reports should be addressed to:

INSTITUTE OF EDUCATIONAL RESEARCH,
Division of Field Studies (Civic Education),
Teachers College, Columbia University,
NEW YORK CITY, NEW YORK.

Journal of Educational Research

OFFICIAL ORGAN OF THE NATIONAL ASSOCIATION OF DIRECTORS OF EDUCATIONAL RESEARCH

VOL. IV DECEMBER, 1921 No. 5

CONTENTS

Entered as second-class matter at the Post Office at Menasha, Wisconsin, under the act of March 3, 1879.

Journal of Educational Research

Copyright 1921, by
Public School Publishing Co.

Official Organ of the National Association of Directors of Educational Research

ARTICLES TO APPEAR SOON

HINES, HARLAN C.: WHAT LOS ANGELES IS DOING WITH THE RESULTS OF TESTING.

MINER, J. B.: AN AID TO THE ANALYSIS OF VOCATIONAL INTERESTS.

MINOR, RUBY: PRINCIPLES OF METHOD APPLIED IN CONCRETE SITUATIONS.

ROGERS, BERTHA M.
and
BAKER, TERESA: A DIAGNOSTIC AND REMEDIAL ACTIVITY IN SUPERVISION.

BROOKS, SAMUEL S.: SOME USES FOR INTELLIGENCE TESTS.

JOURNAL *of* EDUCATIONAL RESEARCH

VOLUME IV | DECEMBER, 1921 | NUMBER 5

COMPARING THE EFFICIENCY OF SPECIAL TEACHING METHODS BY MEANS OF STANDARDIZED TESTS[1]

SAMUEL S. BROOKS

District Superintendent, Winchester, New Hampshire

In the last article five principal factors in a teacher's efficiency were distinguished—namely, (1) managing ability, (2) natural aptitude, (3) method of teaching, (4) interest and industry, and (5) personality. The position was taken that no one of these five factors can be accurately and objectively measured independently of any or all of the other factors.

Although *method* was one of the factors mentioned we nevertheless now propose to measure the efficiency of methods. Note, however, that we do not propose to do so independently of the other factors.

In general the efficiency of a teacher and the efficiency of her methods are pretty much inseparable. It is a mooted question whether or not there can be a good teacher without good teaching methods. We hear it argued, for example, that a good teacher with a poor method will accomplish more than a poor teacher with a good method. This argument implies that good teachers using poor methods may secure better results than poor teachers using good methods, in the same way that a good carpenter with few and poor tools can do a better job than can the novice with the best and most complete set of tools obtainable. We must admit that there is much truth in the argument. Sometimes we find that a teacher who is ignorant of approved methods but who has great natural ability is obtaining better results than another teacher who is without natural aptitude but who, perhaps with all the advantages of normal training, is using or rather misusing,

[1] This is the seventh article by Superintendent Brooks on the general topic, "Putting Standardized Tests to Practical Use in Rural Schools."

the most approved modern methods. One has the true teaching instinct and ability to apply general principles and the other lacks these advantages.

Whatever may be the actual relations between good and poor teachers and good and poor methods, we can all agree, I think, that the best teachers are those who combine natural aptitude with thorough knowledge of up-to-date methods together with skill in applying them so as to realize their possibilities. And although we cannot measure the efficiency of a teacher's methods entirely apart from consideration of her general ability, there is a way, nevertheless, by which we can with the help of standardized tests obtain fairly accurate comparisons of the efficiency of various special methods, taking at the same time full cognizance of the teacher's general ability.

This can be done somewhat as we solve simultaneous equations in algebra—that is, by manipulating the various quantities so as to eliminate all but one of the unknowns. The value of the remaining unknown is readily found after the others are equalized so as to cancel each other. Yet it cannot be said that the eliminated quantities are ignored. The manipulations required to bring about the conditions suitable for their elimination give them their full force in evaluating the result.

And so, if we are to find the relative values of two or more special teaching methods, we must equalize as far as possible the conditions under which those methods are tried out, thus eliminating all the unknown quantities but one. The chief of these external conditions that would affect the accuracy of our results are the general ability of the teachers, the average mental abilities of the several groups of pupils, and the time devoted to class work with the method.

Now, there are two ways in which we may want to compare methods. We may want to discover which of two or more special methods of teaching a subject will give the best results when used by teachers of equal general ability or we may want to learn which of two or more special methods can be used to best advantage by a certain teacher.

To illustrate the first case, suppose we wanted to compare the results of drill in the fundamental operations of arithmetic as conducted in the usual more or less unorganized manner and without much regard for the special difficulties involved in definite types of

examples, with results of drill in the same operations by means of the Courtis Standard Practice Tests. To do this we should first choose our teachers for the trial. Their general ability should be as nearly equal as possible in order to eliminate so far as may be any inaccuracy in our conclusions due to differences in ability. Two teachers with approximately equal ratings by the method described in my last article would serve admirably. One should have had no experience with, and if possible no knowledge of, the Courtis Standard Practice Tests or of similar practice material, while the other should have had experience in their use and knowledge of their basic principles. It would not do to have the same teacher try to handle both methods because, on the one hand, if she had had experience with the practice tests, the defects of the haphazard procedure would be largely nullified by her knowledge of the principles underlying them; while, on the other hand, if she did not have such knowledge and experience, the advantages of the Courtis method would in some measure be lost.

The next step is to choose two groups of pupils. These groups should be neither too large nor too small; neither large enough to be cumbersome to handle as a class nor small enough to make average scores meaningless. From ten or twelve to twenty in a group is probably about right. The pupils in both groups should be in the same grade, and the average mental ages and average intelligence quotients of the two groups should be as nearly equal as possible. The pupils' mental ages and intelligence quotients are obtained of course by means of intelligence tests, some uses of which will be discussed in a later article.

As soon as the pupils have been selected, they should be carefully tested by means of standardized tests in the fundamentals, and their scores should be recorded. The testing of both groups and the scoring of the papers should be done by the same person, preferably a person experienced in such work. The period of drill should begin as soon as the tests have been given. Care should be taken to see that, in each group, exactly the same amount of time is devoted to drill in the fundamental operations each day. At the end of eight or ten weeks the tests should be given again, the scores recorded, and the progress of the two groups compared. The difference in progress of the two groups will approximate the difference in efficiency of the two methods.

The degree of accuracy of the results will depend upon the care with which the tests are given and the degree to which the conditions of the drill work are equalized. It is an open question whether or not the teachers themselves should be informed of the main purpose in view—that is, the purpose of comparing the efficiency of the two methods. If we could be perfectly sure that both teachers would be thoroughly interested and honest about the experiment it would undoubtedly be wise to seek their intelligent cooperation, since by so doing we should be more likely to get the best possible results from both methods. But if, thinking that their reputations are at stake, one or both are likely to be tempted to stretch the time limit for daily drill or to persuade the pupils to drill themselves for speed and accuracy outside of class, then it will probably be better to leave them in blissful ignorance of the main plot, merely seeing to it that each teacher devotes the same amount of time to class drill in the fundamentals each day. In this way one can infer what each of the methods would accomplish under everyday working conditions in the hands of equally competent teachers. If one is particularly desirous of getting the best results of which either method is capable, this purpose may perhaps be accomplished by asking each teacher separately to do her very best.

This particular problem was worked out in my district last year with rather interesting and fairly conclusive results. The Courtis Standard Practice Tests were not in use in the district but, wishing to introduce them the following September, I planned ahead to have the stage set for their appearance. That is, before the practice tests were introduced generally, I wanted if possible to prove definitely that better results could be accomplished by their use with less drudgery for both teachers and pupils.

This was before the teachers' ratings had been computed so I did not have their ratings for guides in selecting the teachers to carry out the experiment. But I did have the records of progress for each school as shown by the September and February tests. Wishing to secure as accurate results as possible under the circumstances, I tried the experiment in each of three different towns. To handle the work with the practice tests three teachers were selected (one in each town) who had shown interest and capability in adapting new ideas to classroom use and whose schools had made normal progress during the first half of the year. Five

weeks before the end of the winter term these three teachers were furnished with the Courtis Standard Practice Tests, Teacher's Manuals, and Students' Practice Pads. I showed them how to use the tests, pointed out their advantages, and explained the principles underlying them. Then I told them that for special reasons of which they would be informed in due time, I was anxious to have them become as expert as possible in using the tests by the end of that term. They assured me that they would do their best and I believe they did. At any rate, they did exceedingly well.

The other three teachers, one in each town, were chosen because their schools had also shown about normal progress for the first half of the year, and because of the further fact that they were all teachers of many years' experience, somewhat set in their ways and not taking kindly to new ideas, but withal hardworking, trustworthy, and capable of doing very good work in their own ways. In other words, they were good old-fashioned teachers.

The intelligence tests had been given by this time throughout the district and I hastened to get the mental ages and intelligence quotients of all the pupils for use in selecting the several groups. They were finally selected according to the plan outlined above except that the grades in any one school were too small to allow of groups of ten pupils being selected from the same grade in such a way that the six groups would all average the same in both mental ages and I. Q.'s. However, the conditions regarding M. A.'s and I. Q.'s were strictly observed and allowed for. The lowest mental age in any group was 10 years, 9 months and the highest was 11 years, 5 months. The I.Q.'s ranged from 97 to 105.

Using the Woody scales for measuring the ability of pupils in the fundamental operations, I gave the first tests to the six picked groups during the first week of the spring term, and corrected and scored them myself, tabulating the average scores for each group in each subject as shown in Table I, in the columns marked "A." At the time I gave the tests to each group of pupils, I had a talk with their teacher, telling her that for very important reasons I wanted her to see how much improvement she could bring about in that particular group during the ensuing twelve weeks by drilling the pupils together just fifteen minutes each day for speed and accuracy in the fundamental operations of arithmetic. The

TABLE I. AVERAGE SCORES IN THE WOODY SCALES

(a) Groups not Using Practice Tests

Operation	Group 1		Group 2		Group 3		Averages	
	A	B	A	B	A	B	A	B
Addition	11.6	14.8	12.0	14.7	11.8	15.1	11.8	14.8
Subtraction	8.2	10.6	8.4	10.3	7.9	10.5	8.2	10.5
Multiplication	8.5	12.2	8.3	12.0	8.1	12.4	8.3	12.2
Division	5.5	8.5	5.9	8.0	6.1	8.8	5.8	8.4
Mixed fundamentals	21.0	25.9	21.0	26.1	22.2	27.0	21.4	26.3

(b) Groups Using Practice Tests

Operation	Group 4		Group 5		Group 6		Averages	
	A	B	A	B	A	B	A	B
Addition	11.9	16.0	11.7	15.8	11.7	16.3	11.8	16.0
Subtraction	8.1	12.4	8.6	12.2	8.0	11.9	8.2	12.2
Multiplication	8.0	15.3	9.0	15.5	8.4	14.8	8.4	15.2
Division	5.4	9.6	5.7	10.2	5.8	9.3	5.6	9.7
Mixed fundamentals	22.0	29.0	19.5	30.0	23.0	29.6	21.5	29.5

three teachers trained for the purpose were directed to use only the Courtis Standard Practice Tests for the drill, but to use them for all they were worth. None of the teachers had any inkling of the real object in view. Yet each one was keyed up to do her best after her own fashion. Every pupil in the six groups was promised a special holiday for not missing more than one day during the term. Pedagogically, of course, this may have been questionable, but psychologically it proved very effective; and I hoped that the end would justify the means. At any rate, I know that a large majority of the pupils got their holiday.

The work was supervised as closely as possible throughout the term. Neither from observation nor by questioning pupils could I detect any evidence that the rules of the game were being disregarded by any of the teachers. At the end of twelve weeks the pupils were again tested with the Woody scales. The average scores for each group were placed in the "B" columns of Table I in such a way that each group's second score in each subject was opposite its first score in the same subject. According to the

table, the average score of the pupils of Group 1 on the first test in addition (Column A) was 11.6. The score for the same group in the second addition test was 14.8 as shown in the first "B" column. The scores for the three groups which did not use the practice tests were averaged for both first and second tests; and they are recorded in the fourth "A" and "B" columns, while in like manner the general averages for the three groups which used the practice tests are recorded in the last two columns of the table.

It will be noted that, according to the "A" columns of the general averages, the two main groups, the first consisting of the three smaller groups in which the practice tests were not used, and the second, of the three groups in which they were used, started almost exactly even in the race as might be expected under the circumstances. The first score of both groups in addition was 11.8 and the first score in subtraction for each group was 8.2. The remaining first scores differed by but one or two tenths of a unit. But this correspondence is no longer apparent when the "B" columns of general averages are considered. The final scores of the group using the practice tests are seen to be considerably larger than those of the group not using them. The differences between the scores contained in the fourth and last "B" columns represent the difference in progress of the two main groups.

The group of pupils drilled with the practice tests has all the best of the argument, the difference in progress being sufficiently great to prove conclusively considerable superiority for the Courtis method properly handled. On the whole, the improvement of all the groups was surprisingly large for a period of only twelve weeks. It amounted on the average to about a year of progress for the groups which did not use the practice tests and to about a year and three-quarters for the group using the practice tests. This merely goes to show what can be accomplished by intensive work along definite lines when the interest of teachers and pupils has been thoroughly aroused.

Now to return to the second way in which we might want to compare special methods. Suppose we wish to learn which of two or three special methods will give the best results with a particular teacher. This is quite a different matter from measuring the relative efficiency of the methods themselves. Only in exceptional cases can methods be accurately compared when handled by the same teacher. For such a purpose the teacher must be equally skilled in the use of the methods to be compared

and without prejudice in favor of any particular method. In particular she must have a thorough knowledge of the special advantages and disadvantages of each method and know how to minimize the latter and make the most of the former. In no other way can the methods be given a fair trial. Only an exceptionally well-trained and widely-experienced teacher, with the impartial mind of a scientist seeking truth through experiment, can fulfill these conditions. Such teachers are not to be found in every school system.

We know that quite often a method of teaching which has proved highly successful when handled by its originator or by teachers specially trained by him, has failed miserably when introduced into an alien school system where the teachers were trained and experienced in other methods. And such failure is not to be wondered at. When the mere form of a new method, without its spirit, is introduced among workers lacking a knowledge of the proper technic to accompany the method, and naturally prejudiced in favor of their own methods, the new method is foredoomed to failure. A few of the better teachers, specially endowed with adaptability and initiative, may grasp the essential advantages of the new method, gradually evolve a suitable technic to fit it, and adopt it as their own. But most teachers, finding themselves accomplishing less with the new method than they did with the old, and longing for the familiar routine, will, unless constant supervision prevents, return surreptitiously at least to their former procedure, convinced that there is none better and that attempting new methods is a waste of time and trying to the nerves.

Of course if the real interest of the teachers can be aroused in the new method by a judicious advertising campaign before it is introduced, and if everybody's patience holds out long enough, and a definite policy of teacher-selecting and teacher-training is carried on, eventually the new method will come into its own if it really possesses marked advantages. But in too many instances the innovation is discarded as worthless after a few months of half-hearted trial without any adequate attempt having been made to modify the environment to fit the new method. And the chief factors contributing to such failures in attempting to introduce new methods of teaching into a school system are the indifference of teachers or their actual antagonism toward new methods, in general, their lack of knowledge concerning

particular new methods, and their lack of foresight and initiative in adapting themselves and their ideas to changing conditions. Probably the most annoying factor and the one most difficult to eliminate is the teacher's mental attitude toward new ways of doing things, her clinging to familiar trails, and her aversion to breaking new paths. even in the interest of finding a smoother, shorter, and pleasanter road to her goal.

Hence new methods, unless real interest and belief in them have been aroused in the teacher beforehand, have to contend against ignorance and indifference or prejudice from the start. I repeat, therefore, that the efficiency of new methods cannot be accurately compared with that of old methods if the new ones are tested by the very teacher whose own methods are being questioned as to their comparative worth. Her attitude is too much like that of the hen defending her chickens from the hawk that would destroy them, the teacher's chickens being her own familiar methods while the hawk is the superintendent with his disturbing new ideas.

We can, however, determine pretty accurately which of two methods a teacher can (or will) handle most efficiently regardless of the actual possibilities inherent in the two methods. And since it is essential that each teacher use, in general, the methods with which she can produce the best results, it is also essential that we know what those methods are. It will not be found profitable, merely for the sake of having certain new methods, to enforce their continued use on teachers who cannot or will not produce as good results with them as they produce with their own methods. So we must have some way of determining whether or not teachers are doing as well or better with the new methods after using them a reasonable length of time, say six months or a year.

This can be done with the help of standardized tests. First select ten or a dozen pupils in the school with mental ages and I.Q.'s as nearly equal as it is possible to arrange. Divide them into two equal groups that average about the same in mental ability. Next test them with some standardized tests in the subject for which special methods are to be compared. Then have the teacher try out two methods, one on each group of pupils, over a period of three or four months. At the end of that time give the tests again and compare the progress of the two groups.

Such a trial will not necessarily prove which method is the better as regards possibilities, nor with which method the teacher *could* do better if she had the proper inclination and training, but it will prove with which method she *will* do better under existing conditions. And that is the essential point. If, after preparing carefully for the introduction of a new method of teaching some subject, by discussing its possibilities with teachers individually and collectively, and by furnishing them with suitable reading material concerning its basic principles, special advantages, and technic; if after demonstrating to the teachers the proper handling of the method and giving them a reasonable length of time to acquire skill in its use; and if after striving in every way to arouse their interest and hearty cooperation in giving the new method a thorough tryout; if after doing all these things and as many more as you can think of, you make such a comparison as outlined above and find that a teacher either cannot or will not do at least as good work with the new method as she did with the old, then it is time to discard either the teacher or the method. If your best teachers have succeeded in getting superior results by using the new method, it means that the method is all right and it may be wise to keep the method and get a new teacher. But if your best teachers have failed to get better results with the new method after several months of earnest effort, it will probably be better to discard the method.

At any rate, in order that the children may get the most for their time and the taxpayers the most for their money, it behooves us to make sure that the methods in use in the schools under our direction are the most efficient that can be used under existing circumstances. We can do this either by selecting and training teachers to fit our chosen methods or by selecting methods to fit the available teachers. Most emphatically it is not efficiency to cling to new methods forced upon untrained or improperly trained and often unwilling teachers just because they are up-to-date methods, when those teachers are not doing as good work with them as with their own methods. Unless we can train our teachers successfully in the proper use of the new methods, or obtain teachers already trained in their use, we had better stick to the old a little longer. Standardized tests will help to prove whether or not the new methods are more successful than the old methods in a particular environment. Results are more important than methods.

UNRELIABILITY OF INDIVIDUAL SCORES IN MENTAL MEASUREMENTS

JOHN L. STENQUIST
Bureau of Reference, Research, and Statistics, New York City Public Schools

With the rapid adoption of tests, including both measurements of intelligence and of educational achievement for purposes of better classification of pupils, a fundamental fallacy has in a large number of cases crept into the statistical procedure, the fallacy, namely, of neglecting to take account of the variability of individual scores.

Many users of tests have innocently passed from the former considerations of grade or school *averages* to the consideration of the status of *individual pupils* without making provision for the unreliability of a single measurement. Undoubtedly the status of the individual pupil is precisely what must be emphasized by school administrators in order to obtain any great benefit from the use of the tests. But caution must be exercised in taking this step lest we fall into serious blunders and reach unsound conclusions.

AVERAGE VS. INDIVIDUAL SCORES

When dealing with measurements of a group—say a class with an ordinary test of intelligence—we customarily compute the class average. The reliability of this average will generally be sufficiently high for all practical purposes. On the basis of it we are justified in inferring that Class A is brighter in general than Class B, that it is less bright than Class C, etc. If a second form of the same test is given, the average derived from it will not differ greatly from the average for the first form. Such class scores are fairly reliable because they are based upon a comparatively large number of measurements. In any such group measurement we really have as many samples of that group as there are individuals. The unreliability of each score is, for the most part, due to chance errors; and these discount each other in the long run and can, therefore, be neglected so far as the group is concerned.

When, however, we turn from the consideration of the score for the group to the consideration of the score of any individual pupil in the group, as determined by a single test, the trust-

worthiness of our score is entirely changed. In a second form of the same test, the score of a given pupil will not necessarily be closely similar to his score in the first form. This is due in part to lack of sufficient sampling of his ability. Many factors come in to change his score; thus he may be physically less fit during one test than during the other; he may be emotionally disturbed at one time and not at another; the precise tasks required in different forms of the same test may for accidental reasons be of unlike dfficulty *for him*, even though they are identical, as shown by group averages. Hence we have the net effect of variations about an average performance. This phenomenon is probably as well known as any other in the whole field of mental measuring; yet, it is a common practice totally to disregard it. In attempting to classify individual pupils by test results, as is now being done in many cities, radical changes in grading are frequently recommended, entirely neglecting the unreliability due to insufficient sampling of each pupil's ability.

The Amount of Variation in Successive Tests

The actual variations which may and do occur in a pupil's score in successive tests of the same trait are a serious matter when we attempt to place him in a specific grade according to any one of such scores. To show concretely what happens, the results for a sample class taken at random out of some sixty or seventy classes in public school 64, Manhattan, is cited in Table I. The tests were given under closely parallel conditions during a period of about four weeks. They were all administered by the same teacher, who was one of the most skillful of forty or fifty who gave tests. All scoring was carefully verified by a recheck on every pupil. An examination of Table I shows the mental age as determined by each test, and the average age according to all five tests (Column 12). The average age is the most significant as it represents five distinct trials on the part of each pupil. Using this as the true mental age, the extent to which each pupil deviated from it in five trials is shown in columns 3, 5, 7, 9, and 11.

Table I reads: Pupil No. 1 obtained a mental age of 166 months according to the National Intelligence Test, Scale A. This was two months less than his mental age as obtained by averaging his results on all the tests he took; his mental age according to the National Intelligence Test, Scale B, was in excess of 180 months or (at least) twelve months more than his mental age as obtained by averaging his results on all the tests he took; etc.

TABLE I. RESULTS IN FIVE INTELLIGENCE TESTS GIVEN BY THE ONE EXAMINER TO ONE CLASS

Pupil No.	National Scale A	Deviation from Average	National Scale B	Deviation from Average	Otis Group Intelligence Scale	Deviation from Average	Haggerty Intelligence Delta 2	Deviation from Average	Thorndike Reading Scale	Deviation from Average	Average Age	Average Deviation
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
1.	166	− 2	180+	+12	179	+11	146	−22	162	− 6	168	10.6
2.	175	− 1	180+	+ 4	182	+ 6	180	+ 4	162	−14	176	5.8
3.	180+	+ 9	180+	+ 9	158	−13	178	+ 7	159	−12	171	10
4.	152	−10	162	0	162	0	163	+ 1	170	+ 8	162	3.8
5.	160	− 5	169	+ 4	160	− 5	173	+ 8	162	− 3	165	5
6.	161	− 4	180+	+15	156	− 9	164	− 1	163	− 2	165	6.2
7.	181	+ 1	180+	0	176	− 4	180+	0	180	0	180	1.0
8.	134	−10	159	+15	150	+ 6	132	−12	146	+ 2	144	9.0
9.	151	−15			165	− 1	174	+ 8	175	+ 9	166	8.2
10.	155	− 8	143	−20	170	+ 7	181	+18	174	+11	163	12.8
11.	156	+13	150	+ 7	138	− 5			127	−16	143	10.2
12.	140	−17	182	+25	152	− 5			155	− 2	157	12.2
13.	160	+ 9	148	− 3	120	−31	155	+ 4	172	+21	151	13.6
14.			147	+17	123	− 7	125	− 5	126	− 4	130	8.2
15.	159	0	169	+10	152	− 7	172	+13	142	−17	159	9.4
16.	160	− 6	180+	+14	163	− 3	170	+ 4	157	− 9	166	7.2
17.	138	−25	180+	+17	176	+13	174	+11	134	−29	163	19.0
18.	159	− 8	154	−13			177	+10	179	+12	167	10.6
19.	180	− 4	180+	− 4	206	+22	180	− 4	175	− 9	184	8.6
20.	150	− 6	160	+ 4			141	−15	172	+16	156	8.2
21.			181	+ 4	178	+ 1	175	− 2	172	− 5	177	3.0
22.			180+	+13	167	0	161	− 6	160	− 7	167	6.5
23.	174	− 3	183	+ 6	180	+ 3	180	+ 3	166	−11	177	5.2
24.	170	− 3	180+	+ 7			182	+ 9	180	+ 7	173	6.5
25.	126	−30	185	+29	154	− 2	157	+ 1	157	+ 1	156	8.5
26.	139	−13	161	+ 9	144	− 8	152	0	164	+12	152	8.4
27.	140	−16	157	+ 1	186	+30	162	+ 6	137	−19	156	14.8
28.	142	−18	180+	+20	146	−14	157	− 3	176	+16	160	14.2
29.	147	−12	180+	+21	141	−18	173	+14	151	− 8	159	13.8
30.	139	−25	180+	+16	182	+18	167	+ 3	154	−10	164	14.4
31.	165	− 9	172	− 2			180+	+ 6	180+	+ 6	174	5.7
32.	180	+ 5	180+	+ 5			175	0	166	− 9	175	4.7
33.	147	+ 6	160	+19	132	− 9	124	−17	143	+ 2	141	10.6
Totals	4686	293	5462	345	4498	258	5110	217	5298	315	5367	295.9
Averages	156.2	9.76	170.6	10.7	160.6	9.2	164.8	7.0	160.5	9.6	162.6	9.0

Five Parallel Tests of General Intelligence

While other tests which are often utilized as intelligence measures were given, they are not here included because of their lower correlations with our criterion. This criterion is the composite (equally weighted) score in the following six tests:

1. The National Intelligence Test, Scale A.
2. The National Intelligence Test, Scale B.
3. The Haggerty Intelligence Test, Delta 2.
4. The Otis Intelligence Scale, Advanced.
5. Meyers Mental Measure.
6. Kelley-Trabue Language.

The coefficients of correlation for each of the tests in Table I with this criterion are:

National A r..... $=0.801$ ($n=560$)
National B r..... $=0.788$ ($n=518$)
Otis (Advanced) $r=0.680$ ($n=551$)
Haggerty, Delta 2 $r=0.808$ ($n=532$)
Visual Vocabulary $r=0.680$ ($n=461$)

In addition to our records in the tests given in Table I, we also have records in the Kelley-Trabue Language tests, in the Woody-McCall Arithmetic, and in Meyers Mental Measure. These, however, correlate with a coefficient of only 0.58 ($n=581$), 0.39 ($n=298$), and 0.48 ($n=544$) respectively, with our criterion. National A, and B, and Haggerty are very similar in nature, and the Otis test does not differ greatly. Thorndike's Visual Vocabulary Test while purely a test of word knowledge, correlates exactly the same as the more elaborate Otis test. The spurious self-correlations do not appreciably operate to the disadvantage of any one test.

Equating Norms

In turning raw scores into mental ages (to secure a common unit) as has here been done, the question of the comparability of norms as furnished by the authors of the tests is involved. These were equated for all but the Visual Vocabulary Test[1] as follows: Twenty-seven classes including 1,007 pupils were measured by all of the tests; utilizing the published norms every raw score

[1] The norms for this test were established with great care and careful statistical procedure by Mr. R. H. Franzen, Director of Educational Research, Des Moines, Iowa, on the basis of some ten thousand cases in New York City and vicinity.

was then turned into a mental age. These twenty-seven classes included six eighth-grade, seven seventh-grade, ten sixth-grade, and four fifth-grade classes.

Considering these twenty-seven classes as a unit, each test should, if it measures the same thing as the others (as it does approximately), and if the norms are comparable, give the same mental age to this unit of 1,007 cases. This we did not find to be the case. Comparing the average age according to each of the tests with the average age according to the Haggerty Delta 2, we found that the average age according to the National A was 13 months higher, and that the corresponding figure for National B was 7.5 months higher, for the Otis 3.7 months higher, and for Myers Mental Measure 6.4 months lower.

The published norms in current usage at the time of our testing and the amount of deviation from Haggerty's Norms (which were taken as a point of reference, because of the large number of cases and their wide geographical distribution) are given in Table II.

When we say that the norms for the National Intelligence Test, Scale A, were thirteen months higher, we mean that the

TABLE II. PUBLISHED NORMS USED

Test	Age										Deviation of Norms from Haggerty Delta 2
	6	7	8	9	10	11	12	13	14	15	
National A			74	79	89	102	113	123	134		13 months higher
National B			76	81	89	102	112	124	128		7.5 months higher
Haggerty, Delta 2			25	43	55	66	77	87	100	115	
Otis Advanced			40	52	64	76	88	100			3.7 months higher
Meyers Mental Measure	9	17	23	28	34	39	43	47	49	49	6.4 months lower

critical score corresponding to each mental age, as given in the manual accompanying the test, was apparently too high, and that on the average it represented the children as being thirteen months younger mentally than they are represented to be by our revision. Accordingly, we lowered the norms for this test, by this amount thus making them less severe. In effect, this means that

for each child when tested by the National Intelligence Test, Scale A, we added thirteen months to the mental age which would have been obtained according to the author's norms. Similarly seven months were added to the ages of chi.dren for the National Intelligence Test, Scale B, and four months for the Otis Scale. In Table I these adjustments have been made.

Analysis of Variation in Five Tests by the Sample Class

With these facts in mind we may examine more in detail the scores of each pupil in Table I.

Column 13 shows the average deviation (in months of mental age) for each pupil. Twelve pupils out of thirty-three or about 35 percent deviated on the average in the five tests, ten months or more, which for practical purposes is about equivalent to one year. If we examine individual cases we find still more serious evidence of how precarious a procedure it is to place a pupil precisely by a single test. Note Pupil No. 12, for example. If his score in National A, had been our sole basis he would have been classified as forty-two months younger mentally than he is shown to be by National B, and seventeen months younger than the average of five tests. Again, note Pupil No. 28, who e deviations in months of mental age are in order: −18, +20, −14, −3 and +16. It will make an enormous difference to this pupil whether he is classified by National A or by National B the difference in his mental age by these two tests being thirty-eight months, or over three years! Almost the same is true for pupils 17, 27, 29 and 30. Six extreme cases are also graphically shown in Figure 1. Here it is apparent at a glance that the variation is very great. A pupil by any one test may easily be misplaced by two years.

The Extreme Cases

The cases selected for Figure 1 are, to be sure, the most extreme ones. It is true that most of the cases deviate less. The average of the average deviations for the class is 8.7 months. But when we propose reclassification of pupils, we propose classifying *all of them.* All are tested, all are placed. But if as many as a third of them may be approximately a year older or younger mentally than a single test result shows; and if a few (here we have at least six out of thirty-five or some 15 to 20 percent), out of every class may be misplaced as much as three or more years, then our work is too unreliable to be utilized for classification purposes,

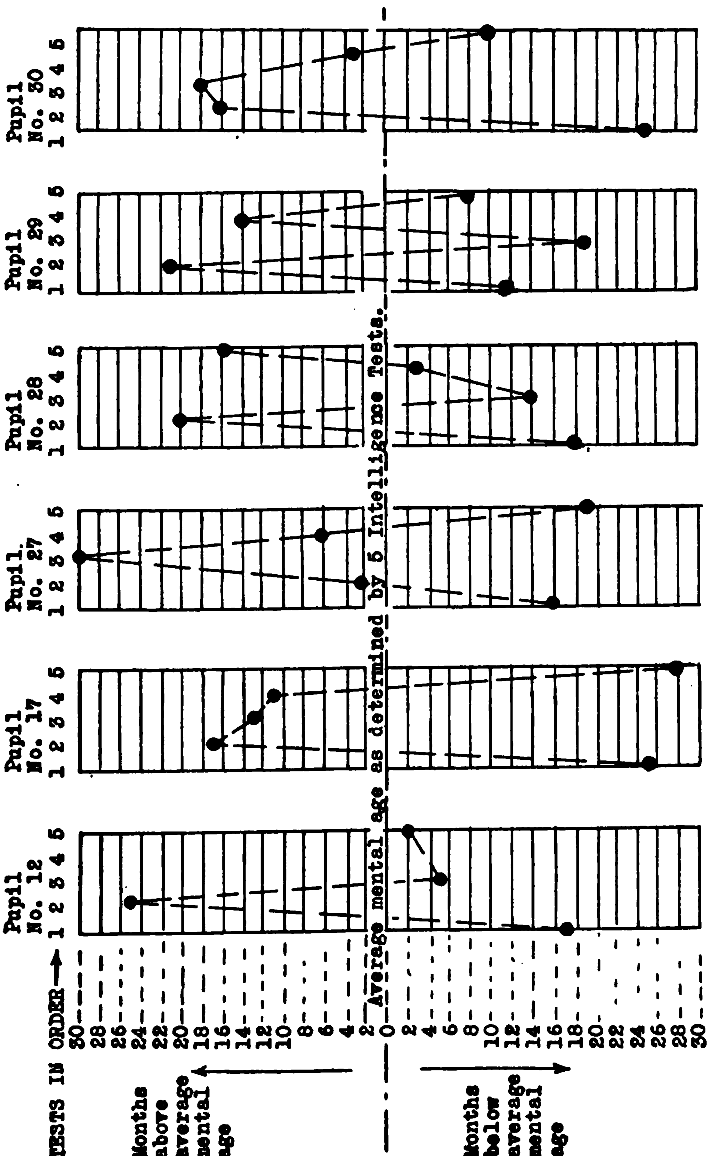


FIGURE 1. SHOWING THE WIDE DIFFERENCES IN MENTAL AGE AS DETERMINED BY 5 SUCCESSIVE INTELLIGENCE TESTS. WHICH IS NEAREST THE TRUTH?

and may be no improvement over ordinary teacher-judgment methods. We believe that variability as great as that we have shown is not unusual. Those who doubt this should provide checks upon their results as we have done and first demonstrate that the great variability that we have found is the exception. It is here maintained that it is the rule; and that unless in utilizing tests to diagnose individuals, measures are taken to overcome it, the present testing movement will fall into disrepute, because tests misapplied cannot do what is claimed for them by enthusiasts.

In connection with this failure of successive test results to agree, we might appropriately raise the question: "Which score shall we accept as most significant?" We commonly assume that the average is nearest the truth. However, as Thorndike has pointed out on several occasions, in nearly all physical feats, we take not the average performance but the maximum. We are not interested, for instance, in the average height which a pole vaulter can clear, but the maximum height. Similarly it would not be unreasonable to use the maximum performance in mental feats. However, this is beside the main issue, namely, the variation in score from test to test.

The Remedy

The remedy is clear, and simple, though to be sure it is not one to elicit great enthusiasm. What is needed is a more thorough testing of each individual pupil—more tests, with equated norms, utilizing of course more and more reliable ones as they become available, and a more and more precise technic. When our results for an individual pupil are even approximately as reliable as our averages for classes then they will withstand any criticism and will be trustworthy. But this means more labor, more expert direction, more test material, and greater cost.

The very least that may be done is to supplement the single thirty-minute intelligence test, now so popular, by at least one other—better still by two or three—before extravagant reports are made concerning the mental status of individual pupils. To repeat: In the past when our inferences were confined to *averages*, we escaped these invalid conclusions. Now that we propose to speak of the *individual*, we must not overlook the statistical implications of the step that has been taken, but provide a technic that will make possible what we are attempting.

THE EFFECT OF KINAESTHETIC FACTORS IN THE DEVELOPMENT OF WORD RECOGNITION IN THE CASE OF NON-READERS[1]

GRACE M. FERNALD AND HELEN KELLER
University of California, Southern Branch

The cases reported in this paper are all those of children of normal mentality who have failed to learn to read after three or more years in the public schools. In all cases but one the vision was normal. The method described here was used only after the child had been given several weeks of individual instruction by recognized methods and had failed to make any improvement.

Many children who have been brought to us as non-readers with individual instruction and proper motivation, learned to read quite easily by ordinary methods when they were given individual instruction and proper motivation; others proved to be mentally deficient. In five years we have found only seven cases of actual non-readers, even though children have been brought to us from all parts of the state. In all seven cases the presumption of mental deficiency had been made as the explanation of the reading failure. In all but one case, however, the intelligence quotient was found to be at least 100 by the Stanford Revision.

METHOD

1. *Learning first words.*—The child was asked to tell some word he would like to learn. The word was written in large script on the blackboard or with crayola on cardboard. The child looked at the word, saying it over to himself and tracing it if he wished to do so. The tracing was done with the first two fingers of the right hand (or of the left hand if the child was left-handed) resting on the copy. It was never done in the air or with pencil. When the child was sure he knew the word, the copy was erased and he attempted to write the word, *saying the syllables to himself* as he wrote them. If he was unable to write the word correctly, the entire process was repeated until the word could be written without the copy. At no stage of the performance was he allowed to copy the word. After a few words had been learned in this way,

[1] This paper is to be followed shortly by two others, giving the results of experiments with first grade children and of experiments in spelling. The bibliography will be published with the final paper.

he was shown the word in print as well as in script. The next day he was shown the word in print only. If he failed to recognize it, it was written for him. If he still failed to recognize it, it was retaught as on the first presentation.

2. *Spontaneous sentences.*—After the first few days the child began to ask for sentences instead of words. A sentence was then written and he learned the words comprising it, finally writing the entire sentence as many times as he wished—always from memory, never from copy.

The sentences the child had requested were then printed on cardboard or were typewritten. These sentences and others, made of the same words, were read by the child. The same words were repeated in different sentences from day to day.

3. *Words in context or story selected by the child.*—As soon as the child was able to make out simple sentences, he was taken to the library and allowed to select a book. The first paragraphs read were worked over in the following manner. Before the reading, each word which had not already been learned was exposed through an adjustable slit in a piece of cardboard. If the child failed to read the word it was pronounced for him. He pronounced and then wrote the word (as before without looking at the copy). If he had difficulty in writing the word after seeing it in print, it was written for him and taught from the script as in the case of the first words.

4. *Apperception of phrases.*—After the words in the new paragraph had been taken up in this manner, brief exposures of the words were given until the child was sure of them. When recognition was immediate for every word, the slit was adjusted to phrases, and flash exposures of the various phrases were given. The exposures were never long enough to permit the phrases to be read word by word. As many successive exposures as were necessary for recognition were given. After the entire paragraph had been worked over in this way, the child was told to read the paragraph to himself and report what he had read.

5. *Silent reading for content.*—As soon as possible the child was encouraged to read to himself. There was no difficulty in any of our cases in getting him to do this after his progress had gone into the fourth phase as described above.

DESCRIPTION OF CASES

Seven cases have been successfully treated to date. Four of these are described in the following pages. In one case, included under the above statement, the work is incomplete at the present time, but the results showed normal progress during the period of experimentation. Two additional cases, one from New York, and one from Colorado, now under investigation, have both reached the third stage. In the latter of these two cases the child is now able to learn new words rapidly from print and to write them correctly although he was unable to read or write his own name two months ago.

CASE I. BOY (LESTER)

Age (December, 1916) 10-2; mental age (Binet-Simon, 1911 Scale), 9-0. Re-tested December, 1920; age, 14-2; mental age (Stanford Revision), 13-4; I. Q., 94. Vision, normal.

School history.—Lester had been in the public schools of Los Angeles for five years (including one year in the kindergarten). When brought to our notice he had been in the second grade of the Normal Training School for two months, having been entered in the Training School as a last resort after his failure in the city schools. On account of his size, he was first placed in the second grade; but, as he was unable to distinguish one-syllable sight words, he was put back into the first grade. Here a special effort was made to teach him to read in a small group where special instruction was possible. Since he made no progress whatever, he was reported to the psychology department as mentally deficient.

As the results of the mental test did not show sufficient retardation to account for his complete inability to read, we kept the boy under observation for some time. The case came to our notice at a time when we were particularly engaged with poor spellers. Accordingly, we put Lester into a group of our worst spellers, where individual work was being done.

Method.—The work in the spelling class referred to was being conducted in the following manner: The children watched while the word was written on the blackboard or on cardboard. They said the word over to themselves, tracing it with their fingers if they wished to do so. The tracing was done with the first two fingers of the right (or left) hand in contact with the copy. Tracing in the air or with pencil did not seem to produce the same results as tracing with the fingers. When each child was sure of the word, it was erased or otherwise removed from view, and the child wrote the word, saying the syllables to himself as he wrote them. He was allowed to write the word as many times as he wished, provided he did not copy it from a previous writing. Except for this special work in spelling, the method was the same as that described on page 355.

Results.—Lester developed a craze for writing words. He would work at it by the hour, tracing words over and over again. To our surprise,

after the first week, he seemed to be able to write from memory words which he had written previously several days in succession.

We had the same words which he had learned to write printed; and each time the boy wrote a word in script we showed him the corresponding printed form. Recognition was developed almost immediately for the words he had written. At first he learned only two or three words a week, then several a day. After about two months, he was able to look at a word in print, say it over to himself, and write it correctly. This was true even of words of three and four syllables. At this stage of his development, after he had once written a word, he would almost invariably recognize it on successive presentations. Yet, on the other hand, if told a word over and over again on successive days, he failed to recognize it unless he wrote it.

Since the boy was interested in history, and especially in the war, we chose a history of the United States as our first book. The method employed is described in the introductory statement on method.

Unfortunately we had no idea of the significance of the work and kept no records of the exact words written from day to day, as we did in our later cases. Much of the writing was done on the blackboard and immediately erased.

The boy seemed suddenly to go ahead by jumps. At the end of five months he was reading so readily that we had a demonstration before his various first- and second-grade teachers. They refused to believe that he was actually reading until they brought books selected by themselves and tried him out with both oral and silent reading.

After six months in the University Training School, the family moved and the boy went back to the regular city schools.

December 15, 1920, just four years after our first experience with Lester, we found him doing satisfactory work in the *seventh* grade of the city schools. His teacher was surprised when we asked about his reading, and said he was an inveterate reader.

TABLE I. RESULTS OF READING TESTS

Date	Test	Results					
Dec. 1916	Several	Zero on all tests; could not read or write words like "cat"					
Dec. 1920	Kansas	Lester's Results		Standard		Ability Shown	
	Silent, FormII, Test 2	Rate 106	Comprehension 24.7	Rate 106 (For Eighth Grade)	Comprehension 7th Grade, 23 8th Grade, 26.4	Rate 8th Grade	Comprehension Above 7th Grade

Note: Lester had never seen any of the upper grade Kansas Silent Reading Tests before the test was given December 1920.

Summary of case. At the age of ten years, after over four years in the elementary grades of the public schools, Lester was totally unable to read.

In six months he developed from a zero score in reading ability to ready word recognition. For the last three and a half years, without further special instruction, his development in both silent and oral reading has been perfectly normal. During this period he has advanced from the first to the seventh grade.

CASE II. BOY (HENRY)

Age, 9-8; Mental Age (Stanford Revision), 12-8; I. Q., 131. Vision, normal.

School history.—Henry entered the kindergarten of the University Training School when five years old, and has attended the same school continuously to date, except for part of one year lost on account of measles and whooping cough.

He seemed unable to learn to read or spell although he had been given individual instruction, including drill in phonics. He was referred to the psychology department as a possible mental defective. Although the results of the mental tests indicated unusual ability, he was found to be below second-grade rating in the Kansas Silent Reading Test and in the Ayres Spelling Scale. He also failed completely in all tests for phonics.

Method.—The method was the same as that previously described, except that Henry did not tend to trace the words. He would look at a word which had been written for him, say it over slowly, have the copy removed, and then write it saying each syllable as he wrote it. He soon developed considerable skill in writing words in this way; but he was very slow at first in making the association between the word he had written and the printed word. In the beginning it was necessary to present each word daily for five or six days before the association with the printed word was formed.

There was so little progress for the first few weeks that the psychology department was consulted repeatedly with reference to the seeming failure of the boy to develop word recognition. His mother came often to observe and expressed herself as convinced that it was simply another failure. She felt that she had tried everything. Besides being in a small class with a particularly successful teacher, he had received special help at home and special work in phonics with the reading department. Every effort had been made to encourage the boy; he had even been promoted to the fourth grade in order to try the effect of encouragement. As he could do nothing with fourth-grade reading and spelling he went back into the special study room. It was with difficulty that we obtained the mother's consent to have the boy tested mentally. She has since told us that she was afraid he would be found mentally deficient. When informed of the results of the test, she was frankly skeptical. It should be stated, however, that she was very careful not to let the boy know of her discouragement and that she cooperated with us throughout the experiment.

After the fifth week Henry suddenly began to make such rapid progress that it was almost impossible to keep track of him. He could write any word after a single presentation and after he had once written it, he would recognize it in either script or print on subsequent presentation. As one of the student teachers expressed it, he was "on." Five weeks after our first work

with him, he wrote the word *Switzerland* after one presentation and recognized it in print twenty-four hours later.

Henry's first reading was in the second reader. It was necessary to write the new words and to have him write them before he could recognize the word on subsequent presentations. The development of a reading vocabulary was very rapid, however, after the first few weeks, since it was rarely necessary for him to write a word more than once, and since he was able to give one glance at the printed word, say it, and write it. He worked much by himself, asking anyone to pronounce a word, and then writing it off on a slip of paper.

From the time when Henry began working alone it was difficult to keep track of his progress. By May, three and a half months after we began the experiment, he was working out many new words without having them pronounced for him or without needing to write them. Though he had received no instruction in phonics since the experiment began, and had seemed at that time to have no knowledge of them, a test given May seventeenth showed the beginning of the development of the recognition of phonetic units.

Work was begun in February and discontinued in June at the close of the school year. The mother reported that the boy read everything available during the summer—library books, newspapers, advertisements, etc. On his return to school he went into the regular fifth grade. At the date of this writing (March, 1921), he is doing satisfactory work in all his subjects. Formal tests given in November showed his ability to be between that of the fourth and fifth grade according to both the Starch and Kansas tests, and above that of the sixth grade according to the Gray Oral Reading Test.

Formal reading, spelling, and phonics tests.—The following results were obtained from tests of spelling, reading, and phonetics. On January 30, 1920, Henry was given a selection of words from Column I of the Ayres Spelling Scale. His rating was ten percent, and the character of this performance may be judged from the fact that the second-grade standard for these words is fifty percent. The following are the word forms as he actually wrote them, the words in parentheses being the correct forms.

1. I (catch)	8. gon (gone)	15. bay (buy)
2. lake (black)	9. se (suit)	16. spo (stop)
3. wrme (warm)	10. trak (track)	17. wa (walk)
4. inles (unless)	11. w (watch)	18. grat (grant)
5. Cothing (clothing)	12. daas (dash)	19. soke (soak)
6. be (began)	13. fell (fell)	20. nu (news)
7. ababl (able)	14. fight (fight)	

On April 25, 1920, another selection of words was given from Column I of the Ayres scale. This time Henry's record was ninety-five percent which compared favorably with the standard for the fifth grade (ninety-four percent).

On February 6, 1920, and again on November 22, 1920, Henry was tested with the Kansas Silent Reading Test, Form I, Test I, with the following results. In the first test his rate score was 36 and his comprehension

score was 4. The lowest grade for which standards on the Kansas Silent Reading Test are available is the third grade. For this grade the standard in rate is 52 and in comprehension, 7.2. Thus, Henry's performance was considerably below that which one would expect of a typical third-grade child. On the second application of the test (November, 1920) Henry's rate score was 103 and his comprehension score was 18. This compares favorably with the fifth-grade standard of 89 for rate and 19 for comprehension.

On November 22, 1920, Henry took the Starch reading test for the third, fourth, and fifth grades. His scores in comparison with the standards for each of these grades follow.

Test	Henry's Score		Standard	
	Speed	Comprehension	Speed	Comprehension
3rd grade	2.3	33	2.1	20
4th grade	3.2	31	2.4	24
5th grade	2.6	21	2.8	33

Phonics test.—Although every effort has been made to teach Henry phonics in the second and third grades, the results of all phonics tests in January, 1920, were entirely negative. He had no idea of the sound significance of letter combinations.

The following results were obtained in tests given by the reading department. No instruction in phonics had been given since the experiment was started..

Test Given May 17th, 1920

1. *ask*—bash, hash, mash, crash, flash, nash
2. *in*—fin, sin, bin, lin, tin, men
3. *ed*—bed, led, he
4. *et*—let, set
5. *ind*—blind, sind, find, dind
6. *ong*—bing
7. *end*—fend, send, senll, mend
8. *ase*—base, mose
9. *am*—fan, sam, fan
10. *ile*—
11. *all*—fall, ball, mall, hall, tall, dall
12. *inch*—pinch
13. *ale*—pale, fale, sale, male, tale
14. *age*—page, cage, mage
15. *ad*—sad, mad, lad

Test Given December 24th, 1920

1. *hack*—rack, tack, mack
2. *all*—ball, tall, call, mall
3. *ate*—hate, mate, fate, late
4. *ent*—rent, ment, sent, tent
5. *ing*—ring, sting, bing
6. *ow*—wow, cow, row
7. *air*—hair, lair, rain
8. *ail*—mail, lail, fail, rail
9. *and*—hand, sand, cand
10. *ang*—rand, dang, sang
11. *ank*—sank, bank, flank, tank
12. *ark*—mark, hark, lark
13. *atch*—catch, match, atach
14. *eck*—reck, heck
15. *end*—attend, blend, afend
16. *ure*—sure, picture

Summary of spelling and reading tests.—(a) Spelling: Progress from below second grade to fifth-grade standing in three months. (b) Reading: Progress from much less than third-grade record to fifth-grade in six months. (c) Phonics: Although no instruction had been given in phonics between January and May, and although Henry seemed to have no knowledge of phonics at the earlier date, tests given at the later date showed the development of associations between many letter combinations and their sound values. Tests given in December of the same year showed his knowledge of phonics to be quite equal to that of the average child in the fifth grade.

Summary of case.—A year ago, Henry was not able to do second-grade work in reading and spelling. After four months' instruction, he was put into the fifth grade and has had no difficulty in keeping up to the fifth-grade standards for the five months of this year. He is now in the second half of the fifth grade, and is reported as reading books of every description out of school hours.

CASE III. BOY (FRED)

Age, 9-2; mental age (Stanford Revision), 9-3; I. Q., 100. Vision: right, normal; left, two-thirds normal.

School history.—Fred attended the public schools of Riverside, California, where he entered the kindergarten at the age of five. He spent three years in the first grade. He entered the University Training School, September, 1920, and was placed in the second grade but made no progress during the first month. He was unable to read words of one syllable, and consequently was too poor to be tested by any formal test. He was sent to the Psychology Department for a mental test with the definite suggestion that he be transferred to one of the public school rooms for the mentally deficient. The child had so much the appearance of a mental defective that we were surprised at the results of the mental tests.

A month after his entrance into the University Training School Fred was given tests in phonics and in spelling. In phonics he wrote such unrelated forms as the following: for *on*, *am;* for *at*, *oot;* for *in*, *tn;* for *ish*, *ik*. In spelling he missed all but two of eight words, writing the following forms for the words given in parentheses:

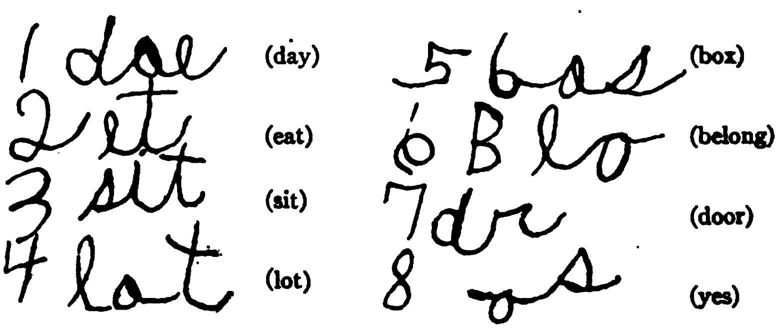


FIGURE 1. WORDS WRITTEN BY FRED SEPTEMBER 29, 1921

First phase—learning first words. Fred spent the morning learning to write the four words *will*, *you*, *am*, *boy*, all of which he had asked for. The words were taught one at a time; each being written by the teacher on the board. Fred said each word over and traced it with his fingers until he was sure he could write it. It was then erased, and he attempted to write it. If the first attempt was not successful, the entire process was repeated until the word was written correctly.

Fred succeeded in writing each of the words correctly in the course of the morning. In the afternoon he attempted to rewrite the same words with the following results: *W* (for will); *Yo* (for you); *ow* (for am); *b* (for boy). The words were retaught and then written correctly. He then asked for the word "I," and wrote:

I am o boy

He wrote this sentence with his left hand, then with his right. There is little difference in the performance with the two hands, but Fred says he likes the left hand better. He usually writes on paper with his left hand and on the board with his right.

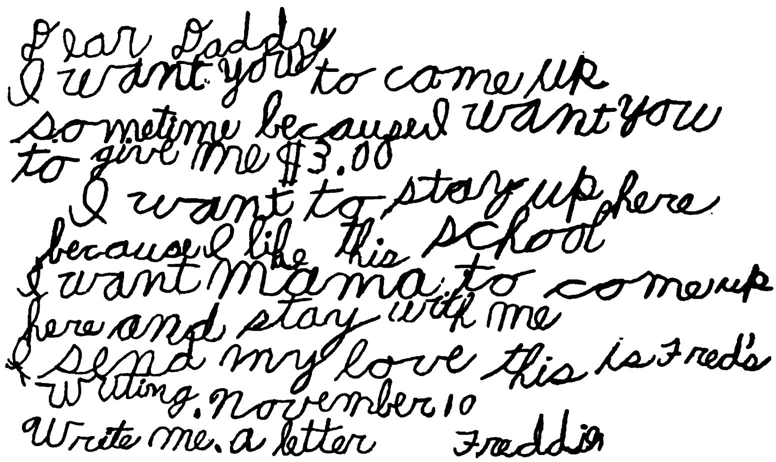
TABLE II. RECORD OF FIRST WORDS LEARNED BY FRED

Date (October)	Word	Number of Times Presented	Date (October)	Word	Number of Times Presented
6	will	2	7		1
7		0	7	yes	2
6	you	5	8		1
7		2	25		0
8		1	8	last	3
25		0	10		2
6	am	3	11	little	1
7		1	12		0
8		0	11	Pine	2
6	boy	1	12		0
7		0	11	tree	1
7	door	2	13		0
7	box	3	11	grew	5
7	open	1	13		0
7	mouse	4	11	woods	5
7	and	2	13		0
8		2	12	October	2
8		2	13		0
25		0	12	of	1
7	we	1	13		0
7	day	3	12	words	2
7		1	21		0
7		0			

October seventh.—He wrote correctly from memory the four words which he had learned the previous day, then asked for *door* and *box*. After working for some time on these words, he asked for *open*, adding "then I'll learn *the* next and I can write *open the door*." When the word *open* was written by the teacher, Fred said, "Erase it quick so I can write it." He next asked for the sentence *Open the box till the mouse jumps out*. The word *mouse* had to be presented four times before Fred could write it correctly. By "presented" is meant that the word was written by the teacher and traced by the pupil until the pupil felt sure he could write it.

Table II gives the record of the first words learned by Fred and the number of times it was necessary for each to be presented before he was able to write it correctly. Table II also shows the results of attempts to write the same words on later dates. The words were all asked for by Fred.

For the next three weeks Fred went through an orgy of writing. He wrote blackboards full of words, then began to write sentences, and finally wrote the letter to his father, as shown in Figure 2.



Dear Daddy
I want you to come up
sometime because I want you
to give me $3.00
I want to stay up here
because I like this school
I want Mama to come up
here and stay with me
I send my love this is Fred's
Writing. November 10
Write me a letter Freddie

FIGURE 2. LETTER WRITTEN BY FRED WITHOUT ASSISTANCE NOVEMBER 10, 1921 (THE WORDS IN THE LETTER HAD BEEN PREVIOUSLY ASKED FOR AND LEARNED)

This letter is only one specimen of the spontaneous compositions with which the boy kept himself occupied by the hour. It was written six weeks after the attempt to write the words given in Figure 1. He was constantly asking for new words which he learned and wrote from memory. For over a month after writing his first sentence, Fred seemed to care little about subject matter, so absorbed was he in mastering the mechanics of writing. He worked so constantly and so hard that it was necessary to force him to leave the room at recess and at the close of school. He looked up one day after working for two hours at new words and exclaimed, "You know I

scarcely ever used to get promoted and now just look at all I am learning." As soon as we attempted to teach him words without having him write them, his interest was lost. He would try to learn them by saying them over and looking at them, but would soon become discouraged and would fail to recognize the words after repeated presentation. This attitude in the early stage of the experiment is particularly interesting, because at the present time (March, 1921) Fred no longer wishes to write unless he has something special to say, but is reading everything he can get hold of. He is as eager over making out new words without writing them as he ever was over the writing process.

Second phase—The association of written with printed words.—Words for which Fred asked were written for him on the typewriter or were shown him in print after he had written them. Simple reading exercises were given, using the words he had written.

Table III shows the results with the first words studied for the purpose of reading. The printed word was first shown the boy and then written for him. He studied it on first presentation as he did the words in Table II, but he was not asked to write a word a second time unless he failed to recognize it on later presentations. Column 3 shows the number of presentations necessary before the word was written correctly.

TABLE III. RESULTS OF FIRST WORDS PRESENTED FOR READING PURPOSES

Date of first presentation (Oct. 1920)	Word	Number of presentations	Date of first presentation (Oct. 1920)	Word	Number of presentations
14	pussy	2		build	3
	where	3		tired	3
15	fur	4	20	month	3
	would	4		sleep	1
	rings	2		dinner	1
	thanked	3	21	chair	1
	fields	1		bowl	2
	over	1		middle	2
18	sunshine	2		bears	1
19	great	2	25	heard	1
	name	1		across	1

Third Phase—Writing the word from memory after looking at the printed copy and having the word pronounced.—Within six weeks after the experiment was started, it was never necessary for Fred to have the word written for him. He was able to look at printed words of several syllables, say them over to

himself, and write them correctly from memory. On November seventeenth he wrote *disappointed*, *department*, *university*, and *training* after seeing the words once in print. A week later he recognized all of these words without any hesitation, and without having seen them in the interval.

Development of projects in connection with reading and writing. November seventeenth.—Fred suddenly decided that he wanted to draw, and was allowed to do so as long as he wished. He drew a picture of a cannon and then of a house and garage with heating system. His picture of the garage is shown in Figure 3. It was then suggested that he label his pictures, and this idea pleased him greatly. He learned to write the following seven words and used them as labels: *cannon*, *shells*, *pipe*, *steps*, *window*, *garage*, *furnace*.

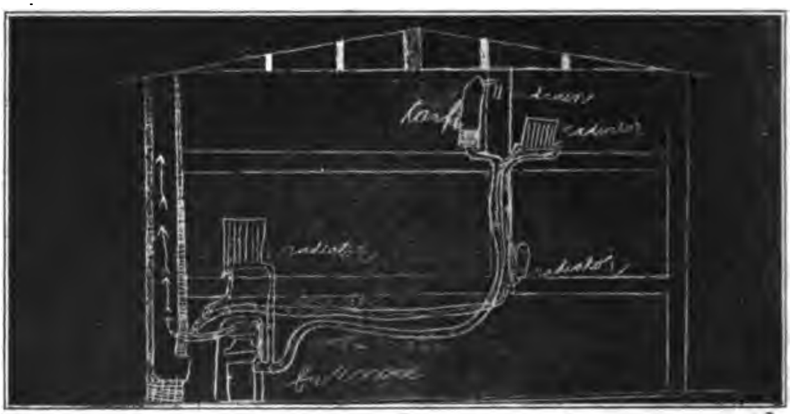


FIGURE 3. PHOTOGRAPH OF ONE OF FRED'S DIAGRAMS

November eighteenth.—He wrote correctly without presentation the following words which he had previously learned: *you*, *ride*, *bicycle*, *furnace*, *hide*, *pipes*, *rope*, *because*, *garage*, *cannon*, *want*, *Riverside*, *December*, *October*, *will*, *live*. He then wrote above the words on the blackboard, "These words are saved."

November nineteenth.—He was taken to the library and allowed to select two books: Lynde's *Physics in the Household*, and a book on plumbing. From November nineteenth till December sixth, Fred drew and labelled diagrams, finding the words in physics and plumbing books. He was still unable to recognize new words, after he had been told what they were, unless he wrote them. One writing even of very difficult words was usually all that was necessary for word recognition.

November twenty-second.—Fred read the following paragraph without being told any words except those enclosed in parentheses:

> The cooler water in the radiator (being heavier) sinks from the radiator into the furnace (boiler) and (forces) the hot water from the (boiler) into the radiator. This hot water gives up its heat to the air in the room and (thus) cools (contracts) and becomes heavier. It then sinks back into the boiler and forces more hot water into the radiator.

This paragraph was read just one and a half months from the time when the boy could not read the simple sentence "*I am a boy.*" He stumbled over a simple word like *thus* which he had never written, but had no difficulty in reading any of the much harder words which he had written. He was told *thus* each time he failed to read it, but was quite unable to recognize it on later presentations until he had written it, after which he recognized it whenever it occurred.

December sixth.—Fred suddenly stopped drawing and labelling diagrams and began writing on the blackboard sentences which he would sometimes illustrate with pictures. These sentences and others, containing the same words, were then typewritten and given him to read. For example, the following sentences with a picture of a racing machine at the top of the sheet were written after a visit to the automobile machine shop.

A RACER

This is a fine automobile.
It has four wheels on it. They have tires and mud guards.
It has a steering gear, and a crank on the front.
It has a windshield. It has a radiator.
It is a Mack and it is a racer.
It runs awfully fast.
Gasoline runs the engine.
The radiator has water in it to cool the engine.
The engine has eight cylinders.
This machine has United States tires.
The engine is a fine one.
The fan helps keep the engine cool.
This machine can turn corners going very fast and it won't wreck.

December 13, 1920.

The following 206 words, written from October sixth to December twenty-second inclusive, are arranged in the order of the frequency with which Fred used them. The numeral after each word indicates the number of times it was used.

On December 22 all the words in this list were shown to him and he recognized all of them except the four marked with an asterik.

I	25	water	11	because	8	me	5
these	22			cannon	6	on	5
a	18	want	10	has	6	Riverside	5
				pipes	6	words	5
and	14	it	9	radiator	6		
this	14					come	4
the	14	cool	8	can	5	draw	4
to	14	furnace	8	day	5	fine	4
		ride	8	engine	5	fur	4
like	13			garage	5	little	4
		hot	7	hide	5	October	4
is	11	will	7	live	5	*page	4

saved....4
tired....4
where....4
would....4
yes....4

bicycle....3
box....3
boy....3
build....3
car....3
chimney....?
crank....3
December....3
door....3
fast....3
*month....3
mouse....3
Mrs....3
rope....3
stay....3
up....3
with....3
write....3

are....2
arithmetic....2
autumn....2
cold....2
have....2
heating....2
in....2
last....2
love....2
machines....2
mama....2
now....2
of....2
open....2
pipe....2
pussy....2
read....2
rings....2
sun....2
saw....2
school....2
send....2
sunshine....2
system....2
tank....2
*thanked....2
tires....2
umbrellas....2
university....2
we....2
wires....2
window....2
wood....2
won't....2
wreck....2

across....1
air....1
an....1
as....1
auto....1
automobile....1
awfully....1
ball....1
bears....1
beat....1
belong....1
birthday....1
blew....1
board....1
boat....1
bought....1
boiler....1
bowl....1
bridge....1
buttoned....1
buy....1
cackle....1
carry....1
chair....1
circus....1
closer....1
coming....1
corners....1
cylinders....1
daddy....1
dear....1
department....1
didn't....1
dinner....1
disappoint....1
drew....1
easy....1
eight....1
expect....1
fan....1
faster....1
fields....1
fireplace....1
four....1
front....1
garden....1
gasoline....1
gear....1
give....1
going....1
grew....1
heard....1
help....1
here....1
holiday....1
house....1
hurt....1
*journey....1
letter....1
lot....1
lazy....1
middle....1
motor....1
mud guards....1
my....1
name....1
never....1
November....1
one....1
over....1
pin....1
pine....1
play....1
plough....1
poles....1
racer....1
railroad....1
riding....1
rose....1
run....1
runner....1
shells....1
shoe....1
sign....1
sit....1
sky....1
sleep....1
sometime....1
spark plug....1
stand....1
steering....1
steps....1
tail....1
than....1
that....1
track....1
training....1
tricks....1
trolley....1
turns....1
unable....1
under....1
United States....1
very....1
walk....1
warmer....1
wheels....1
wire....1
writing....1

All but four of the words in the above list were asked for by Fred and learned because he wished to read or to write about some particular thing. The four exceptions were "department," "disappointed," "university," and "training" which were given to demonstrate the readiness with which he could learn new words.

The numbers after the words indicate the number of separate occasions upon which the word was written in context. The boy often wrote a word over and over again voluntarily, erasing it each time it was written, but as no record was kept of this the actual number of times each word was written is much greater than indicated.

December twenty-second.—Fred began to ask for new words and to remember them without writing them. *December twenty-fourth—January fourth.*—Christmas vacation.

Fourth phase—Ability to pronounce new words if they resemble words already written. January twenty-fifth.—Fred sounded out the word *mother*, then said it over several times. He was quite excited over being able to say the word without being told, and began to attempt the same thing with other words.

February fourth.—He worked out, without writing or any help, the following words: *surprised*, *roar*, *fright*, *noise*, *thirsty*, *dirty*, *middle*. He had to be told *dreadful*, pronounced *towards*, *to-wards*, and *vines*, *vī-nĕs*. He read the fable of the *Fox and the Lion* to himself and told the story, giving every detail. He had never heard or read the story before.

March third.—Words were given to Fred from the pages of the third reader. The longer words on various pages were given out of context to see whether he could read them without having them pronounced. All the words were new to Fred so far as we know. He read all the words in the following list, mispronouncing only those which are starred. In each case the mispronunciation is indicated.

enough	money	roasting	*ŏnly
crowing	with	ashes	wounded
screamed	tripped	quite	howled
tumbled	*wármth	loaded	stolen
midnight	frozen	lazy	empty
*chămber	traveler	healthy	shoulder
*quarreling	chestnut	branches	*cruelly
quarrel	graceful	whining	brooks
mistress	*grĭnned	*răttling	California
maid	monkey	apple	state
*d~~o~~ūbled	second	grave	series

At the date of writing this paper (March 1921), Fred's progress is so rapid that it is difficult to keep records. He takes library books home, reads to himself, and has to be told only such words as might trouble any child of his age. He has been in the regular third grade for two weeks and is having no difficulty with the work. If his progress continues at the present rate, he should be able to make up the one year necessary to put him in the grade appropriate to his chronological age.

Summary of case.—In October 1920, Fred could not read or write even monosyllabic words. He failed completely in all tests for reading, spelling, and phonics. His school report showed steady attendance with failure of promotion in the city schools of Riverside, California. It was taken for granted that he was mentally deficient until mental tests proved him normal.

At the beginning of the experiment his progress was very slow. He seemed wholly dependent on tracing the words first learned and continued to trace for over two months. His development went through well-marked stages, with the transition from one stage to the next apparently quite sudden.

In March 1921, five months after the experiment was started, he was reading and writing (spelling) well enough to go into the regular third grade. He was taking library books home to read to himself and could read any ordinary story and give its content.[3]

CASE VI. (JOE)

Age (June 1919), 12-5; mental age (Stanford Revision), 15; I. Q., 120. Vision, normal.

Although incomplete, this case is reported because it differs in certain respects from our other cases. *In the first place*, in spite of the fact that the boy showed almost no ability in reading or spelling, he did not have to go through the first two stages described on page 355. From the very start he could write a word from memory, after seeing the printed word, having it pronounced for him, and pronouncing it himself. Thus, he was able to begin with what had proved the third stage in all our other cases. *In the second place*, the experiment was discontinued before the fourth stage was reached. Joe had learned many new words so that he could read fairly well on certain topics; the work with rapid apperception of phrases had been begun; but the stage had not been reached in which new words were apperceived on the basis of their resemblance to words already learned. The progress of the case during the year and a half of regular school work since the experiment was discontinued is of interest in comparison with the progress made by cases in which the experiment was completed.

School history.—Joe has attended the Los Angeles city schools since he was six years old. From the start he has seemed to be unable to learn to read but has been passed from grade to grade on account of his ability in other subjects, particularly in arithmetic. His grades in the fifth year ranged from *B* in arithmetic through *C*− in geography and history to *D* in English. When questioned as to how he kept up the reading end of some of his school subjects, he explained that he had always managed to work with a boy who could read easily. He helped the other boy with problems in return for help in reading.

Joe was referred to us in June 1919 as a sixth-grade failure. Investigation of the case showed that the entire difficulty was due to an almost complete inability to read. There had been considerable discussion of the case before it came to the psychology department, since two other children in the same family had had the same difficulty with reading.

The case was the more surprising because of the boy's high intelligence quotient and because of his history outside of school. He excelled in games like baseball and was very popular with other boys. The proprietor of a

[3] At the date of publication, December 1921, Fred is doing satisfactory work in the upper fourth grade.

drug store where he worked out of school hours reported that he could keep track of things better than most men.

Since the case came to our attention at the end of the school year, we transferred the boy to a summer school where the experiment could be started. He attended this school intermittently for a month. Since then he has had a few hours of special work, extending over a period of about two weeks. He entered the seventh grade in the fall and is now (a year and a half later, March 1921) in the eighth grade in one of the regular Los Angeles city schools, where he is reported as doing satisfactory work.

Method.—The method was the same as that already described except that, like Henry, Joe never tended to trace words, and was able from the start to learn them directly from the printed copy. Consequently in the case of each word it was only necessary to pronounce it for him, to have him pronounce it to himself as he looked at it, and then to have him write it from memory, saying each syllable as he wrote it.

Results.—Although the results of the first reading and spelling tests were almost entirely negative, Joe possessed from the beginning a remarkable ability to learn new words by the method just described. In spite of the fact that he was unable to recognize short, ordinary words unless they were pronounced for him, he was able to look at a difficult word, pronounce it after someone, and write it correctly from memory, provided he said each syllable to himself as he wrote it. It seemed to make no difference whether he knew the meaning of the word or not. He could write a word like *psychophysical* after seeing it once and repeating it after some one, and could often rewrite the word, as well as recognize it, after an interval of several days. His ability to write words in this way was so unusual as to attract general attention. It became one form of school entertainment to try to find a word which Joe could not write. To this day he knows certain outlandish words learned in this way. It was, of course, not a part of the plan of instruction to teach any but ordinary words. During the five weeks of summer school he went over the Ayres thousand-word list and developed a considerable reading vocabulary.

A study of Joe as he works shows that he is more dependent than any of our other children on saying the word as he writes it. He makes marked lip movements as he says the syllables to himself and fails completely if he is made to suppress the lip movements.

He usually remembers for an almost indefinite period any word he has ever written, but fails on a subsequent presentation to recognize a word which he has been told repeatedly but which he has not written. At the end of the summer school, when the experiment was discontinued, he had reached the stage where he was beginning to make out a few new words on the basis of their resemblance to words he already knew. For the most part, however, he was still not able to recall words unless he wrote them. Work in rapid apperception of phrases had just been begun.

In the fall of 1919, Joe went into the regular seventh grade and did fairly well in all his studies, except English in which he failed. He seemed to have developed sufficient reading ability to keep up with his work in other subjects.

The experiment was discontinued because his teacher, since he was able to do the work in most of his subjects, considered it unnecessary to take time for the special reading work.

In March 1921, we found Joe in the eighth grade of the regular Los Angeles city schools. He was reported as doing satisfactory work, although slow in subjects requiring reading. Actual tests for reading and spelling give sixth-grade average for spelling and very irregular results for reading. He grades anywhere from the third to the sixth grade by the Kansas Silent Reading Test, below fourth grade by the Starch Silent Reading Test, and makes only half the standard score for the eighth grade by the Gray Oral Reading Test.

The discrepancy in the results of the silent reading tests is easily explained as soon as one observes Joe while he reads. He makes marked lip movements, and stops when he comes to some word he cannot recall, or else mispronounces some word to himself and so loses the sense of the whole. In the latter case he begins again at the beginning and either reads until he gets the meaning or gives up. For example, in the Starch test for the fourth grade, he missed the word *waked*, read till he came to *dawn*, went back to the beginning twice, and finally gave it up after getting *waked* and missing *dawn*. His score for Test 5 was much better than that for Test 4 because he did not happen to stumble over any particular word.

He is still able to learn a new word almost as rapidly as he can look at it and say it to himself. To determine whether the writing of the word is still essential for word recognition, the following experiment was tried:

He wrote the following words from paragraphs 8, 9, and 10 of the Gray Oral Reading Test: *dignifying*, *station*, *position*, *securing*, *approximately*, *scrupulously*, *inclined*, *contemptuous*, *silence*, *complexion*. Although he had failed to read any of these words, he wrote each of them correctly after a single presentation consisting only in the exposure of the word while it was pronounced and repeated once. The two words, *ingratiatingly* and *Josephus*, were the only ones on which he failed.

The words *proportioned*, *exegencies*, *profusion*, and *exhausted* were told to him four times and repeated by him each time, but he was not allowed to write them.

He was then asked to read the paragraphs again. He read correctly each of the words he had written but failed on all of the four words which had been pronounced but not written.

The results of educational tests given both at the beginning of the experiment and in March, 1921, indicate the amount of improvement which Joe made during the interval. These tests covered the subjects of spelling and reading. In spelling, words from Column Q of the Ayres scale were given on June 3, 1919. Joe's rating was 10 percent, which may be compared with the standard rating of 58 percent for the fourth grade on words of this column. On June 9, 1919, words from Column Q of the Ayres scale were again dictated. This time, however, the test was given after the words had been learned. Joe's rating was 90 percent. On March 7, 1921, he was tested with words from Columns Q and S of the Ayres scale. On Column Q

his rating was 85 percent (standard, 84 for the sixth grade); while on Column S his rating was 80 percent (standard, 73 for the sixth grade). In this connection it may be noted that his school report for March 7, 1921, indicated that his spelling was "good."

The reader may judge the amount and character of Joe's difficulties with spelling by consulting the following list of words from Column Q of the Ayres scale as Joe actually wrote these words on June 3, 1919. In each case the word in parenthesis is the word attempted.

somtines	(sometimes)	dclaer	(declare)	ingae	(engage)
factoer	(factory)	finel	(final)	cention	(connection)
troruble	(terrible)	frem	(firm)	serprise	(surprise)
regen	(region)	prived	(period)	convick	(convict)
addtion	(addition)	privet	(private)	imploy	(employ)
proped	(property)	debate	(debate)	sclet	(select)
proin	(crowd)				

Table III shows the results at the beginning and end of the experiment according to the Gray Oral Reading Test. In June, 1919, his score was 7.5; in March, 1921, his score was 28.75.

TABLE III. JOE'S RECORD ON THE GRAY ORAL READING TEST

Paragraph	June 1919			March 1912		
	Time (seconds)	Errors	Score × Value	Time (seconds)	Errors	Score × Value
1	40	5	0	20	0	20
2	40	3	10	20	0	20
3	35	0	20	25	0	20
4	45	6	0	29	0	20
5	80	8	0	25	3	10
6	85	10	0	30	3	10
7	190	8	0	29	2	15
8	......	.....	.....	90	7	0
9	......	.....	.....	90	10	0

In June, 1919, Joe was quite unable to negotiate the Kansas Silent Reading Test for the sixth, seventh, and eighth grades. He was not given Test I (i.e., the test for the third, fourth, and fifth grades). In March, 1921, he was given both Test I and Test II. On the former he secured a rate score of 61 and a comprehension score of 9. On the latter he secured a rate score of 54 and a comprehension score of 13.

In the Starch Silent Reading Test he obtained the following results in March, 1921: on the fourth grade test, 0.45 words per second, comprehension 11; on the fifth grade test, 1.2 words per second, comprehension 19. An explanation of the inversion shown here is given elsewhere.

Summary of Spelling and Reading Tests.

June 1919. *Kansas Silent Reading Test.* Failure on sixth-, seventh- and eighth-grade form.

Gray Oral Reading Test. Score 7.5 in comparison with 49 as sixth-grade standard. Below first grade if computed on basis of first-grade test.

Ayres Spelling Scale, Column Q. Score 10 percent, standard for fourth grade 50 percent.

March 1921. *Kansas Silent Reading Test.* Grade from third to sixth, varying with test forms.

Starch Silent Reading Test. Below fourth, irregular.

Ayres Spelling Scale. Sixth grade.

Summary of case.—In June 1919, Joe was reported as a sixth-grade failure. Investigation showed that the entire difficulty could be traced to failure in reading and writing (spelling). It was found that he could not read ordinary material such as a simple arithmetic problem, though he could easily solve the problem if it were read to him. He failed almost completely in formal reading and in spelling tests. In spite of his inability to read or spell, the boy was able to learn new words with remarkable speed and accuracy by the methods already described.

After six weeks of rather irregular instruction, he had developed a considerable reading vocabulary along certain lines and had been brought up to the sixth-grade standard in spelling. He had improved sufficiently to be able to read his arithmetic problems and to do the work of the seventh grade, though he still read very slowly and stumbled over new words.

During the year and a half since the experiment was discontinued he has made little progress in ability to read to himself. He seems, however, to have lost none of the concrete detail he acquired, and he still manages to read fairly well in those subjects for which he learned a specific reading vocabulary. His spelling continues to be satisfactory because he learns to write new words easily, and is allowed to use the method developed during the experiment. Unlike our other cases in which the experiment was completed, Joe has never shown any tendency to read to himself and so has not acquired speed and ease in reading.

General Conclusions

In all but one[8] of the cases studied, progress seems to have taken place in four distinct phases, as follows:

Learning to write words.—In all cases the children were at first lacking in ability to write words as well as in ability to read. The development of ability to write words is very slow at first. It is necessary for the child either to trace or articulate the word many times while looking at the written copy, and finally to articulate it as he writes it from memory. The need for

[8] Case IV did not go through stages 1 and 2 but began at stage 3.

tracing gradually disappears, but he continues indefinitely to articulate in learning to write a new word.

Associating the written with the printed word.—The child sees the word in print, has it written for him, and then writes it himself, often tracing difficult words before writing them. He soon reaches a point where he can generally recognize a word in print after he has written it. He must still have the word written for him before he is able to write it himself.

Ability to write a new word from memory after looking at the printed copy and repeating the word to himself.—The word must, of course, be pronounced for him before he is able to say it to himself. He is still unable to recognize short, easy words on subsequent presentation if they are taught him in the usual way and if he does not write them. At this stage he will often write from 25 to 50 words a day. He rarely fails to recognize a word after he has once written it.

Ability to pronounce new words if they resemble words he has already learned.—The end of this stage is normal ability to read. The progress at the end is so rapid that it is almost impossible to keep track of the child's development. He seems suddenly to read and is able to enter regular classes in all work involving reading. In the first three cases the children not only developed normal ability to read, but became incessant readers.

Effect of intelligence on method and learning rate.—The method of learning was practically the same with cases of varying degrees of intelligence, except that there was no tendency to trace in the case of the two children with the highest intelligence quotients. In all cases the articulation and the writing of the word seemed essential for developing word recognition. The progress was much more rapid in the cases of better mentality than in those of lower mentality.

Persistance of kinaesthetic factors.—Children who have to trace words in the early learning stages continue to make slight hand and arm movements in attempting to recall difficult words or to learn new words. All the children make marked movements of articulation during the process of learning a new word, even after reading has been well developed.

Acquiring skill in penmanship and phonics.—Although there has been no drill in penmanship or in phonics, the children who

trace words in their early learning stages write a clear, free hand; and all acquire incidentally a good working knowledge of phonics.

In all of our cases any digression which directed the child's attention away from the word itself seemed to confuse him rather than to hasten the learning process. The introduction of phonics, formal penmanship drill, oral spelling, or even spoken directions, during the writing of the word seemed to hamper the learning process.

Individual differences.—The cases studied differed somewhat in the exact kinaesthetic content necessary for the development of word recognition, the difference being in the amount of hand and arm kinaesthetic experience necessary before the word was written. Case IV did not go through the first and second stages as described above. Although, in this case, the word was not traced before it was written and did not have to be presented in script, it was necessary for the child to write the word before he could recognize it.

General significance of results.—It may be well to note here that, although only extreme cases are reported in this paper, the results of certain experiments now under way seem to indicate that these general principles hold true of many cases in which the child simply has *difficulty* in learning to read. It seems that, at least in many of these cases, the progress will become normal if the proper kinaesthetic content is supplied.

Theoretical

Perhaps we can go no further in theory than to say that, in the specific cases studied, lip and hand kinaesthetic elements seem to be the essential link between the visual cue and the various associations which give it word meaning. In other words, it seems to be necessary for the child to develop a certain kinaesthetic background before he can apperceive the visual sensations for which the printed words form the stimulus. Even the associations between the spoken and the printed word seem not to be fixed without the kinaesthetic links.

The motor tendency is still obvious after the children become fluent readers. They seem far to outclass other children in the same grades in their ability to look at new words, say them to themselves, and write them. All of these children still make pronounced lip movements of saying the words (not the letters)

when learning a new word, even after they have reached a point where they never trace a word or speak it aloud. The children who traced in the beginning tend to make arm and hand movements in learning a new word or attempting to recall a difficult one. They are hopelessly confused as soon as they attempt to spell orally or to write a new word without saying it to themselves.

It would seem that the methods of teaching reading have always neglected the kinaesthetic factors, except those which in no way express the word as written or printed. It has been taken for granted that, in the case of all children, the visual cue is adequate to arouse those associations which make this cue stand for word meaning.

AN EXPERIMENT CARRIED ON WITH THE PUPILS OF THE RUSSELL PREVOCATIONAL ROOM

JAMES H. VOORHEES
Principal, Lynch School, Detroit, Michigan

The Prevocational Room of the Russell School consists of about fifty pedagogically retarded boys who have been weeded out of the regular grades through the Psychological Clinic and placed by themselves as a select group. The fact that the elimination of this type of child from the regular grades affords a great relief to the school system is axiomatic, but the purpose of this experiment was to find whether the proposition as it now exists can be defended on any other terms, that is, whether these boys show any growth in academic knowledge? Do they profit by the academic and manual instruction received here to the extent that they are better able to cope with life as they face it? It is not to be expected that any investigation carried on with a single group will speak conclusively for the various classes of this type in the City of Detroit. It is hoped, however, that in view of the growing importance of the problem the present work will be extended by others who will verify, repudiate, or modify the conclusions herein reached.

The present investigation, it will be seen, divides itself naturally into two distinct phases: first, the measurement of growth or gain in knowledge of the academic subjects pursued in the room; second, a follow-up study for the purpose of finding the exact situation of those who had gone from the room to become members of their respective communities. To determine the amount of growth in the subjects pursued by the boys, the following Detroit Standard Tests were given: spelling, arithmetic, geography, writing, and the Trabue Language Scale. The initial tests were given on separate days during the first week in January, 1919, and the final tests were given during March. Figure 1 shows the results of these scores. It is to be noted that these tests were identical in every case and that the period between the initial and the final tests was about two and a half months. The arithmetic test was a test of the four fundamental operations.

A gain of 16.5 percent was made in spelling, and a gain of 6 percent in arithmetic. The class apparently had not profited from the instruction received in the other branches during this

ten-week interval. The loss sustained in city location may have been due to the fact that but little drill had been given on cities during the intervening period. The small variation in the scores made in handwriting (not shown in the figure) language and state location tends to indicate that each test was a fair representation of the individual abilities and that their limit had been reached.

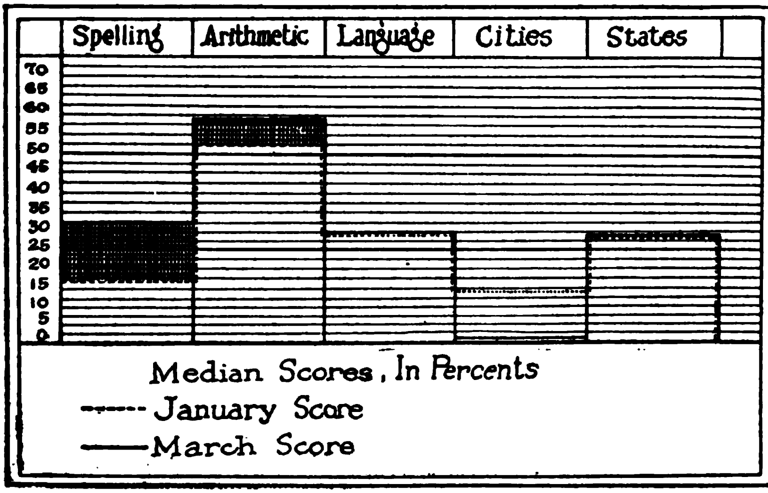


FIGURE 1. GROWTH IN ABILITY MEASURED BY CERTAIN TESTS

A second experiment in academic measurement was carried on during 1919–1920 in the same room. This experiment was an attempt to determine the relative growth in academic knowledge of the Prevocational Class, of a class of the same mental age, and of a class of the same chronological age, all within the same building. A class of B-third-graders was selected as representatives of the same mental age, and pupils of the grade who were at age, neither retarded nor accelerated, were chosen because the average mental age of the prevocational group was nine years. A class of B-eighth-graders who matched these boys in chronological age was also selected. These three groups were given identical tests at the same time in December, 1919, and in May, 1920, the interval being about six months. It was apparent in tabulating the results of these tests that the third graders were the best to use for comparison, the eighth graders having registered too high a score in the initial tests even though the time allowance was adjusted. The

time allowance for the third graders and for the prevocational class was the same in all tests.

This experiment seemed the more interesting because of the longer intervening period between the giving of the tests and because of the fact that the prevocational boys were given a different teacher during five of the six months that the experiment ran. The tests consisted of the following Detroit Standard Tests: Arithmetic Test No. 16, an addition test; Arithmetic Test No. 6, a reasoning test; Test 1B, a sentence organization test; a spelling test, made from the Ayres scale; a writing test, measuring quality and rate; and a silent reading test, measuring rate of reading and index of comprehension. All of these tests were identical except the spelling test, which was devised from the Ayres scale by selecting different words of the same difficulty for the initial and the final tests. The results of the handwriting are shown first. It is well to note here that a program was arranged by the teacher whereby these boys were given a twenty-minute lesson in penmanship two and three times a week for a four months period by the assigned penmanship instructor of the school. The results of this experiment are compared also with the results of the year before when practically no attention was given to penmanship as a subject. Figure 2 shows the scores made in these tests.

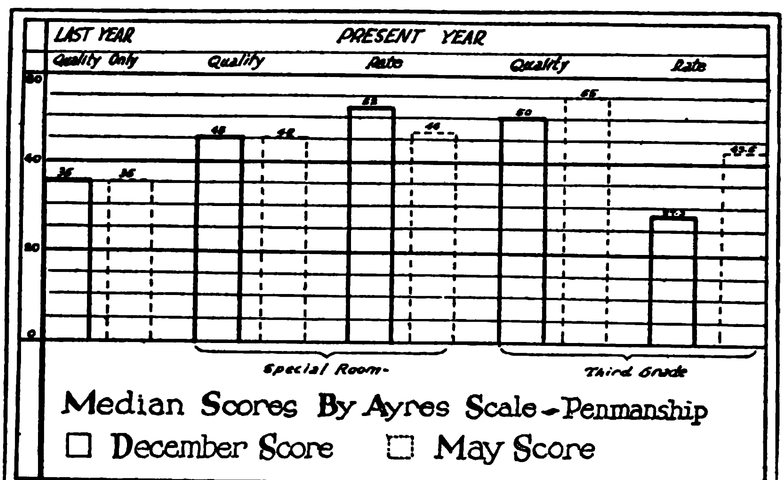


FIGURE 2. COMPARISON OF GROWTH IN PENMANSHIP OF A PREVOCATIONAL AND A THIRD-GRADE GROUP

The same quality of handwriting was shown in each test of the first year, and in the scores of 1919–1920 the same quality was attained in each test while the rate had slowed down six words. Evidently the class had not profited by the four-months period of efficient penmanship instruction. The experiment was very much worth while, however, in view of the added light thrown on the investigation. The third grade over the same period of time showed a gain in quality from fifty to fifty-five, and in rate, from twenty-seven and three-tenths to forty-three and five-tenths words per minute. The quality was ten points higher than that of the prevocational boys, and a good gain was made in rate of writing.

Figures 3 and 4 show the results of the scores made in arithmetic, sentence organization, and spelling by each class.

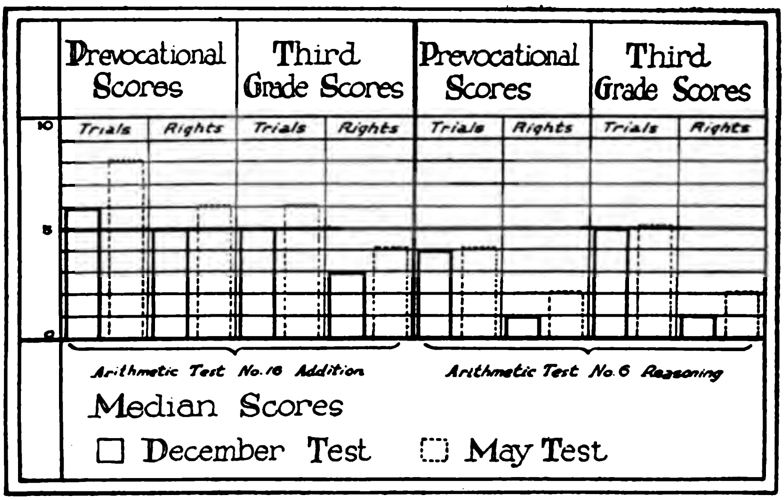


FIGURE 3. COMPARISON OF GROWTH OF A PREVOCATIONAL AND A THIRD-GRADE GROUP MEASURED BY TWO ARITHMETIC TESTS

The outstanding feature of these graphs is the great similarity between the gains of the two groups, and the low scores of the prevocational group. For example, in the addition test the prevocational boys had profited by this six months interval to the extent that they were able to attack two more problems, the third graders one more; the boys were able to solve correctly one more problem in addition, the third graders one more; in the reasoning test both

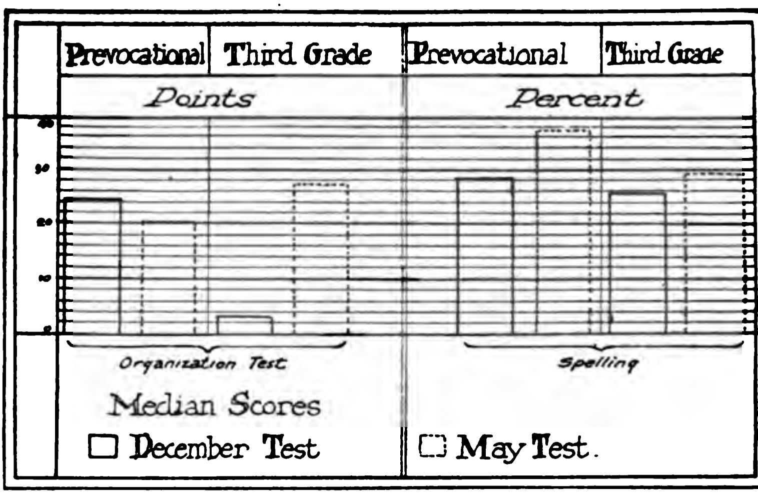


FIGURE 4. COMPARISON OF GROWTH OF A PREVOCATIONAL AND A THIRD-GRADE GROUP MEASURED BY AN ORGANIZATION TEST AND A SPELLING TEST

groups stood the same as far as gains and ability to solve correctly were concerned. In the organization test the prevocational group lost four points while the other gained twenty-four points. The amount of gain here for the third-graders is possibly due to a better understanding of the test after having taken it once, and the last test is probably a better measurement of their abilities than the first one. In spelling the prevocational class seemed to have profited more by their course of instruction over the six-months period, having gained about 9 percent, while the others gained but three percent. However, this is about 7 percent less than the gain made in spelling the previous year. The present test is believed to be a better criterion of their spelling ability, because it consisted of different words, thus avoiding the possibility of any carry over; while in the test of the previous year identical words were used.

Figures 5 and 6 show the scores made in the silent-reading test. Here, too, we observe an equal ability, but a greater gain made by the third-grade group. In the initial test the prevocational boys read on the average 143 words per minute, and 147 words in the final test; the third grade gained from 105 to 145 words, a gain of 44 words per minute, and a rate which exceeded that of the

other group by two words. In ability to interpret the printed page the prevocational boys made a gain of 5.5 percent, the third-graders showed a gain of 17 percent and ranked but 1.5 less than the prevocational group in interpretative ability.

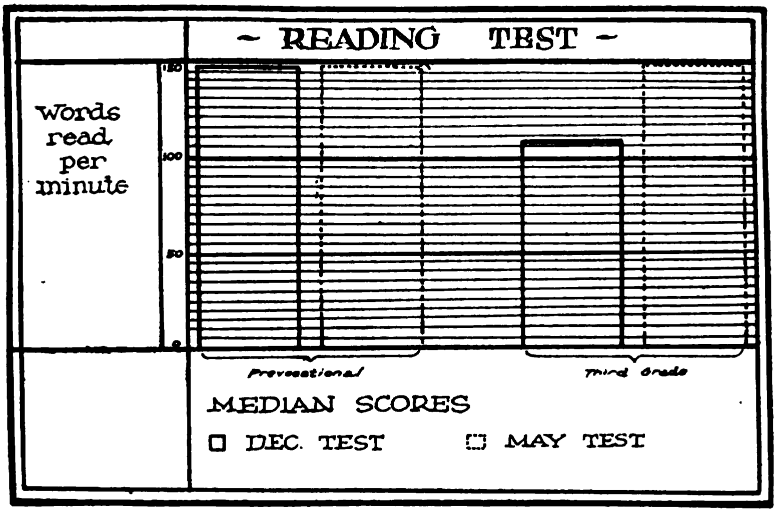


FIGURE 5. COMPARISON OF GROWTH IN RATE OF READING OF A PREVOCATIONAL AND A THIRD-GRADE GROUP

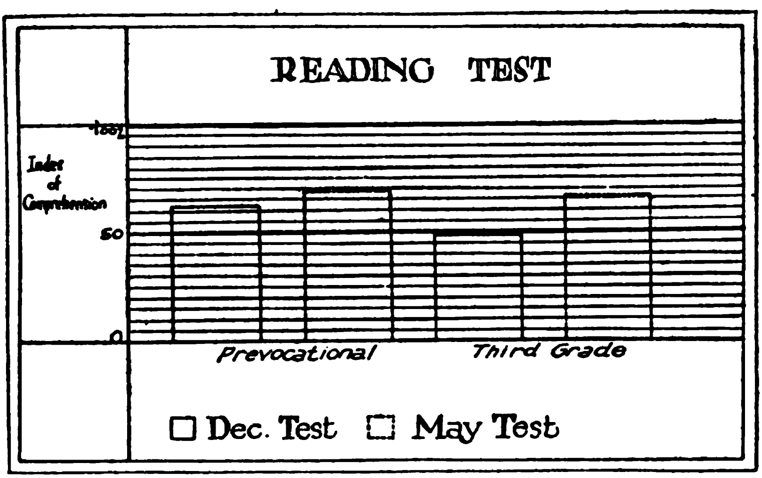


FIGURE 6. COMPARISON OF GROWTH IN READING COMPREHENSION OF A PREVOCATIONAL AND A THIRD-GRADE GROUP

It is quite apparent from the scores made in these tests that the prevocational group has about the average ability of B-third-graders. As a matter of fact, in most of these tests scarcely 20 percent of these boys made scores in excess of the third-grade level, and they did not profit by their course of instruction over this six months period as did the third-grade pupils.

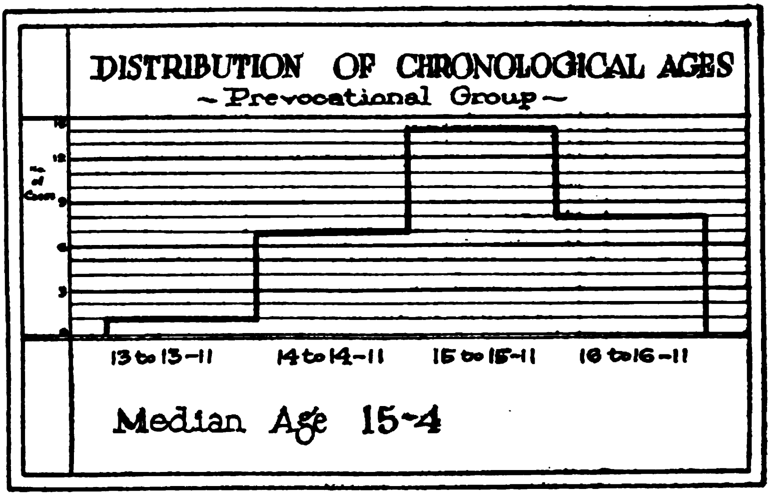


FIGURE 7. DISTRIBUTION OF CHRONOLOGICAL AGES OF THE PRE-VOCATIONAL GROUP UNDER DISCUSSION

Before drawing further conclusions it might be well to turn from this phase of the experiment and get the status of this particular group of boys from another angle. Figure 7 represents the distribution of chronological ages on June 1, 1920. The boys range approximately from fourteen to seventeen years of age, with fifteen years four months as the median. They are, therefore, judging them by the scores made in the tests, on the average, six years behind in school work.

Figure 8 represents the distribution of intelligence quotients. The range here is from fifty-one to seventy-five with sixty-four as the median. Seventy-one percent of these boys had an I. Q. of less than seventy. In view of this fact it is not at all difficult to account for the low scores and little progress made in the academic branches.

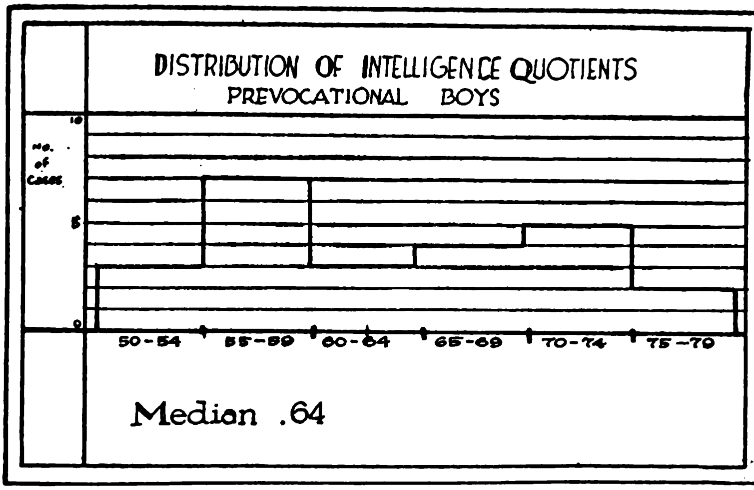


FIGURE 8. DISTRIBUTION OF I. Q.'S OF PREVOCATIONAL BOYS

Figure 9 shows that these boys spent from one to six years, with an average of two years three months, in the special room. This is a sufficient period for them to have profited by both the academic and manual instruction if it were possible for them to do so.

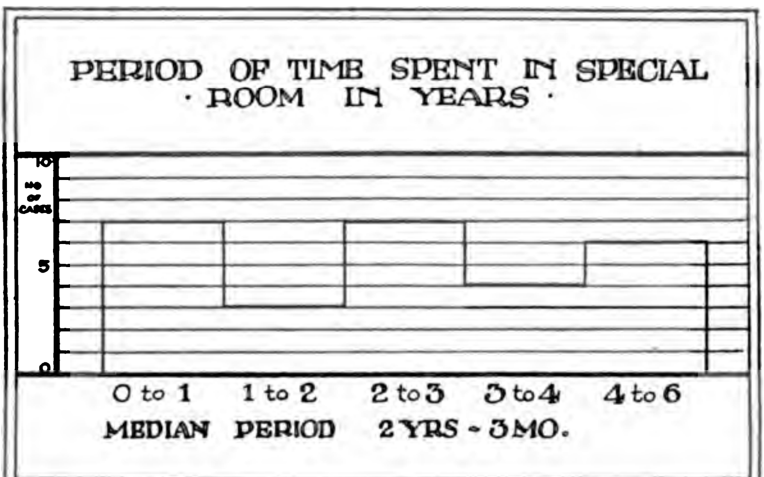


FIGURE 9. DISTRIBUTION OF CASES IN TERMS OF LENGTH OF TIME SPENT IN SPECIAL (PREVOCATIONAL) ROOM

The purpose of the second phase of this investigation was to take an invoice of all the boys who had at some time been members of the Russell Prevocational Class. A list of 125 boys was accordingly made, all of whom had been away from school from one to three years; but, owing to the migratory tendencies of the poorer classes which these boys invariably represent, only 75 were located. Of this number, 27 were found to be unemployed, and 13 of these had not been employed since leaving school while the rest were out of jobs for various reasons. Many had been supplanted by returned soldiers, and possibly many would never have been employed had it not been for the great labor shortage during the war period. Two boys were found in free hospitals, two in the State Institution for Feeble Minded, one in the army, and one in the navy. Two boys had served time in the State Industrial

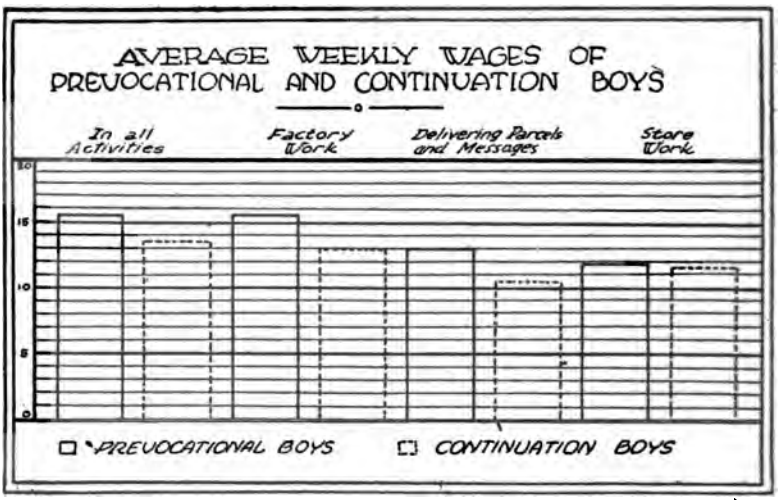


FIGURE 10. COMPARISON OF AVERAGE WEEKLY WAGES OF BOYS ONE TO TWO YEARS OUT OF PREVOCATIONAL SCHOOL AND AN EQUAL NUMBER OF BOYS IN CONTINUATION SCHOOL

School but had been released. Fortunately only one individual had taken upon himself marital duties. He, finding them too strenuous, however, had returned to his parental domicile. Only seven boys were located who had held the same jobs since leaving school. The character of employment and the wages received by these boys seemed to be of vital importance; and in order to make a com-

parative analysis, the average weekly wages of an equal number of boys from the continuation department of the Cass School were checked. Figure 10 shows the results of this part of the investigation. The reader should bear in mind when looking at this graph that all of the prevocational boys were from seventeen to twenty years of age, while the continuation group were all in their sixteenth year. The average wages of the boys from the prevocational class were two dollars per week more than the wages of the continuation boys. A similar situation was found even in the same field of activity as is shown on the right of the figure. In factory work their average excess was two and a half dollars per week; in delivering parcels and messages it was two and a half dollars per week; and in store work twenty-five cents per week.

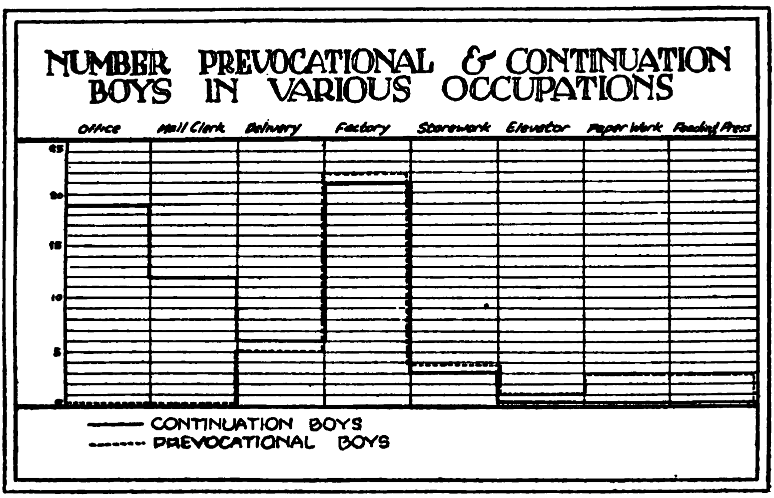


FIGURE 11. DISTRIBUTION OF CASES OF PREVOCATIONAL AND CONTINUATION SCHOOL BOYS BY OCCUPATION

Figure 11 shows the number of boys from both groups employed in various lines of activity throughout the city. The supremacy of the continuation group is here clearly shown, as many of them are found employed in activities for which the prevocational boys are not qualified. It might be added here that of the 75 continuation boys investigated, 57 held the same position throughout the year, and that the eighteen who changed their places of employment apparently did so to better themselves. They were all enrolled in continuation classes, pursuing courses

that had a bearing, in most cases, upon the particular field in which they were employed. On the contrary the prevocational boys seemed to be victims of circumstance with no definite view or purpose in life, getting a job here and there wherever they could find someone to hire them. They seem to be most successful in the factory. No doubt it furnishes the best opportunity for this type of boy because of the better wages and the greater field for stereotyped work requiring a low order of intelligence and initiative. In the factory some of the boys were found to be operating machines, but most of them were doing roustabout work, such as pushing trucks and sweeping floors. No boys were found to be doing wood work or to be making any direct application of the training they had received in the manual training room.

It might be held that an investigation of this kind proves nothing because a repetition a week later would find some of these idle ones employed. However, from the general trend of affairs one would be more likely to find some of the employed ones idle. As a matter of fact, a survey of the 25 boys who left school during the last year, who were not included in the 75 previously recorded, shows similar results. Thirteen of these boys were found to be working, three of whom were not regularly employed; three were selling papers; and only two had remained on the same job since leaving school. One of the twenty-five was deceased, and eleven were known to be idling on the streets.

The conclusions from the investigation seem to the writer to be of lesser importance than the fact that the work was carried on. The inference drawn may be faulty, but the only hope for a more satisfactory solution of this problem as it now stands lies in continued work along these lines. As far as the present survey went, it would seem to challenge the term "prevocational" as applied to this particular room, because the instruction offered there seemed to have no bearing upon the work of these boys after they left. Whether this particular room is a failure in the educational scheme of Detroit depends entirely upon what the room is expected to do. If it is intended for pupils who are retarded and who are later to be reclaimed for the grade, it falls short of its purpose, as only one boy is on record as having returned to his regular grade. If the purpose of the room is to furnish a relief to the regular grades, by eliminating that type of child who does not get on well there, it does all that is expected of it. It is also true that this accomplishment alone is a great piece of constructive work in education. It

would seem, too, that this is about all that can be hoped for under the present order of things. It is generally conceded by those who are best informed that this type of boy will not go very far forward in academic studies, and the present arrangement emphasizes academic instruction. The results of the tests in this experiment are evidences that these boys have not gone very far in academic knowledge, and the I. Q. ratings mean that they have not the possibilities of improving their intellectual status to any great extent.

If any definite conclusion can be drawn from this investigation at all, it is this: that as far as proper returns for efforts expended are concerned, these boys are able to compete with the usual lad far more satisfactorily in the industrial world than they are in the pursuit of academic knowledge. If that be true, it follows that the program for these children should center around the industrial idea. It is not to be inferred that they should be trained specifically for the trades; a general training centering around their probable field of activity is possibly the best that can be done for them. They will be far happier in life if they are able to read, spell, write, do simple arithmetic, understand healthy living, and have regard for law and authority; but outside of these limits they will no doubt get better returns for time and effort expended if their energies are directed along vocational lines.

Certain advantages for a program along these lines can be seen in a centralization plan. For instance, larger groups could be broken up into smaller like units, a departmental plan of teaching effected, a wider range of manual instruction offered, and a placement scheme adopted whereby these boys could be given jobs through the affiliation of the school with the factory or the shop. This scheme could be defended from two angles; first, the apparent adaptability of the children concerned, and second, the economic condition of the parents whom they represent. It might be added here that the average number of children in the families represented by these boys is five, that the father is in most cases a day laborer and a renter, and that every one of the hundred homes visited, required all or part of the boys' daily earnings for their maintenance and support. Many of the circumstances were pressing. If, therefore, a program of this type could be made an actuality, it seems probable that the proposition could be better defended and that the problem of the exceptional child could be more satisfactorily solved for every one concerned.

THE RELIABILITY OF PREDICTION OF PROPORTIONS ON THE BASIS OF RANDOM SAMPLING

Ben D. Wood
Teachers College, Columbia University

The layman and the statistician of little experience find it difficult to accept with any satisfactory degree of confidence predictions based on proportions of comparatively small random samplings. For example, if it is observed in a random sampling consisting of one-quarter of all the sixteen-, seventeen-, and eighteen-year-old boys in a given city, that 83.4 percent have the father as guardian, what would be the proportion of the remaining three-quarters of such boys who would similarly have the male parent as guardian? The average layman would not even attempt to guess within 10 percent of the truth, and he would probably laugh if someone should venture that it would be 83.4 percent plus or minus 2 percent or less. Again, if for the above sampling it were observed that, for 6.3 percent of the boys, the second year high school was the last school grade completed, that for 1.4 percent sickness was the (reported) cause for leaving school, that for 9.8 percent $18 was the (reported) beginning weekly wage, and that 2 percent left school at the age of 13 years, the average person would be far from ready to accept these as anything like the proportions that would be observed in the total group.

Many will welcome the evidence afforded by an empirical study which recently came to light in the form of a test case which is none the less valid for having been made somewhat clandestinely by a group of skeptics. On December 3, 1918, the Vocational Bureau of the New York State Military Training Commission received a questionnaire card from each of the 6,468 employed boys sixteen, seventeen, and eighteen years of age in the city of Buffalo. About 275 public school teachers filled out the cards for the boys. The same thing was done in every part of the state; and in order to avoid the tremendous task of handling so many cards, the present director of the bureau, Mr. H. G. Burdge, upon assuming charge, gave orders that in certain units random samplings be taken which were to be studied in lieu of the total number of cards for such units. The group of subordinates in

TABLE I. RESULTS OF RANDOM SAMPLING[1]

Description of Item	Percent of Enrollment			Description of Item	Percent of Enrollment		
	25	75	100		25	75	100
	1	*2*	*3*		*1*	*2*	*3*
I. GUARDIAN OF BOY				IV. AGE LEAVING SCHOOL			
Father	83.4	82.4	82.4	Ten years or under			
Mother	13.3	14.1	13.9	or no answer	0.8	0.7	0.8
Uncle	0.6	0.6	0.6	Eleven	0.2	0.1	0.2
Aunt	0.4	0.2	0.2	Twelve	0.6	0.5	0.5
Stepfather	0.7	0.9	0.9	Thirteen	2.0	1.9	1.9
Stepmother	0.2	0.1	0.2	Fourteen	31.6	30.1	30.4
Brother	0.5	0.5	0.5	Fifteen	36.9	37.3	37.1
Sister	0.2	0.3	0.4	Sixteen	21.5	23.5	22.9
Headmaster or matron	0	0	0	Seventeen	5.5	5.0	5.2
Grandparents	0	0.1	0.1	Eighteen	0.9	0.9	0.9
Others not related	0.6	0.7	0.7				
No answer	0	0.04	0.02	V. LAST GRADE COM-			
				PLETED			
II. NUMBER OF CHILDREN				Fourth grade or un-			
IN FAMILY				der or no answer	2.1	2.2	2.2
One	6	6.3	6.3	Fifth grade	3.2	3.4	3.4
Two	11.3	11.8	11.7	Sixth grade	14.5	13.5	13.8
Three	14.8	13.7	13.9	Seventh grade	19.7	20.3	20.2
Four	13.6	14.4	14.2	Eighth grade	23.7	26.9	26.1
Five	14.3	14.6	14.5	1st yr. H. S.	23.8	20.4	21.2
Six	11.9	12.6	12.4	2nd yr. H. S.	6.3	6.2	6.2
Seven	9.8	10.5	10.3	3rd yr. H.S.	1.7	2.2	2.0
Eight	8.1	7.2	7.4	4th yr. H. S.	1.8	1.4	1.5
Nine	4.2	4.1	4.2	Business school	3.2	3.3	3.3
Ten	3.0	2.7	2.8				
Eleven or more	2.7	2.0	2.2	VI. BEGINNING WEEKLY			
No answer	0	0.04	0.03	WAGE			
				$ 3.00	10.1	8.6	8.9
III. REASON FOR				$ 6.00	17.4	18.0	17.9
LEAVING SCHOOL				$ 9.00	13.8	15.1	14.8
Financial	9.1	10.1	9.9	$12.00	11.2	10.9	10.9
Wanted to work	68.4	69.4	69	$15.00	14.5	14.4	14.4
Sick	1.4	1.2	1.3	$18.00	9.8	9.4	9.5
Graduated	12.2	11.	11.4	$21.00	7.7	7.6	7.6
Miscellaneous	0.6	0.3	0.3	$24.00	5.6	4.7	4.9
Disliked school	8.3	7.9	8.0	$27.00	2.8	3.6	3.4
				More than $27.00	0	0	0
				No answer	7.1	7.7	7.6

[1] Column 1 shows the proportions in a random sampling of 25 percent (1,617 cases) of Buffalo, New York boys 16, 17, and 18 years of age who have the characteristic indicated at the left. Column 2 shows the proportions observed in a sampling of 75 percent (4,851 cases) of the boys, and Column 3 shows the proportions observed in a 100 percent sampling of the boys (6,468 cases). This table shows parallel columns of proportions for only six of the twenty-four items available for use in this study.

charge of the Buffalo cards was so skeptical that some of them determined to test, *sub rosa,* the wisdom of Mr. Burdge's economy.

Accordingly, the 6,468 cards were put into strict alphabetical order, and every fourth card was extracted. The extracted cards, thus comprising 25 percent of the total and representing 1,617 cases, were sorted and tabulated with the Hollerith machines, as Mr. Burdge had directed. The remaining cards, comprising 75 percent of the total (4,851 cases), were run through the machines for similar sorting and tabulation. Finally, all cards were thrown together and the total 6,468 cards were put through the machines. The results were placed in parallel columns as in Table I. The agreement illustrated ought to put an end to heresy. It is noteworthy that even in the items involving small numbers of cards, the proportions in the three groups are almost identical, clearly demonstrating the sagacity of Mr. Burdge's judgment in the matter.

The parallel columns of the larger table from which Table I is taken afford material for comparing the theoretical with the observed or empirical reliability of the percentile method with random sampling. The formula for the standard deviation of a proportion is

$$\sigma_p = \sqrt{\frac{pq}{n}} \quad \ldots\ldots\ldots\ldots (A)$$

in which p indicates the proportion having the characteristic in question, q the proportion not having it, and n the number of events or cases in which the proportions p and q are observed.[1]

We might test the validity of this formula by comparing the index of reliability which it gives with the index actually observed in the data illustrated in Table I. But the form of the data makes it more convenient to test the formula in a slightly altered form, that is, the form which gives the theoretical standard deviation (σ_{D_p}) of the difference of proportions:

$$\sigma_{D_p} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} \quad \ldots\ldots\ldots\ldots (B)$$

This formula is directly derived from formula (A) by means of the equation for the standard deviation of the difference between

[1] The derivation of this formula is given in detail in chapters 13, 14, and 15 of Yule's *Introduction to the theory of statistics*, C. Griffin & Co., London, 1917. See particularly pp. 257 ff., for all the theoretical considerations of this formula and for the conditions of truly random sampling.

corresponding values of two variables. This equation (Yule, *op. cit.*, p. 210 ff.) may be derived as follows:
Let z be the difference between any two corresponding values, x_1 and x_2 of the variables.
That is, $z = x_1 - x_2$.
Squaring both sides of the equation and summing,

$$\Sigma(z^2) = \Sigma(x^2{}_1) + \Sigma(x^2{}_2) - 2\Sigma(x_1x_2)$$

Dividing both members by n,

$$\frac{\Sigma(z^2)}{n} = \frac{\Sigma(x^2{}_1)}{n} + \frac{\Sigma(x^2{}_2)}{n} - \frac{2\Sigma(x_1x_2)}{n}$$

That is, if r is the correlation between x_1 and x_2 and σ^z, σ_1, and σ_2 are the respective standard deviations,

$$\sigma^2{}_z = \sigma^2{}_1 + \sigma^2{}_2 - 2r\sigma_1\sigma_2$$

Assuming that $r = 0$,

$$\sigma^2{}_z = \sigma^2{}_1 + \sigma^2{}_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Now, since the variabl . are proportions,

$\sigma^2{}_z = \sigma^2{}_{Dp}$, where Dp stands for the difference of proportions,

$\sigma^2{}_1 = \sigma^2{}_{p1}$, and

$\sigma^2{}_2 = \sigma^2{}_{p2}$

Substituting these values in (1)

$$\sigma^2{}_{Dp} = \sigma^2{}_{p1} + \sigma^2{}_{p2} = \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}$$

$$\text{and } \sigma_{Dp} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$$

Formula (B) would inspire us with more confidence if we had empirical proof that it conforms to fact. If, for example, we should calculate σ_{Dp} for various ranges of percentiles by the use of the formula, and if we should find that the standard deviation of observed differences of proportions in these ranges approximated fairly well the theoretical σ_{Dp}, then we should feel more secure in using this formula in working with percents. These calculations for theoretical and observed standard deviations of the differences of proportions have been made roughly with results as shown in Table II.

The values in the column headed "Observed S. D." were obtained by distributing within each percentile range indicated the differences between every pair of proportions in columns 1 and 2 of the larger table from which Table I was taken, and then

TABLE II. STANDARD DEVIATION OF THE DIFFERENCES OF PROPORTIONS EMPIRICALLY AND THEORETICALLY DERIVED

Percentile Range	Number of Observed Differences from Each Percentile Range	Observed S. D.	Theoretical S.D. from Formula B
50-65	13	1.78	1.43
65-75	14	2.15	1.316
75-85	29	1.756	1.149
85-90	22	1.288	0.950
90-94	25	1.259	0.778
94-97	30	0.7865	0.596
97-98.5	17	0.3937	0.426
98.5-99.5	27	0.2816	0.252
99.5-99.8	10	0.1948	0.1675
99.8-99.9	12	0.0913	....
99.9-99.97	10	0.0946	....

calculating the standard deviation in the ordinary way except that the mean was assumed to be zero. Thus, in the table of which Table I is a sample there were 13 pairs of proportions within the percentile range 50–65. The differences between these 13 pairs of proportions gave a distribution with a standard deviation of 1.78 (without correction for assumed $M=0$).

The theoretical values of the standard deviation in the last column of Table II were obtained by means of Formula B described above, namely $\sigma_{D_p}=\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}$, in which p_1 equals a given proportion in Column 1 of Table I, q_1 equals $(1-p_1)$, and n_1 equals the number of cases (1,617) involved in Column 1 of Table I; p_2, q_2, and n_2 $(=4851)$ are the corresponding values for Column 2 of Table I. For the purpose of this rough verification of the formula, the midpoint of each percentile range was taken as the value of both p_1 and p_2. Thus, for the range 50–65 (midpoint = 57.5), the theoretical standard deviation of the differences of proportions is:

$$\sigma_{D_p}=\sqrt{\frac{(0.575)\,(0.425)}{1617}+\frac{(0.575)\,(0.425)}{4851}}=1.43$$

It will be noted that the observed standard deviations are larger than the theoretical; but the differences are consistent,

and from a practical viewpoint not large. The differences are due partly to the roughness of the calculations in both columns, partly to the slight inaccuracies involved in carrying the original proportions to one decimal place only, partly to the slight error introduced by assuming that the mean is zero in calculating the observed standard deviations, and largely to the fact that 275 relatively untrained teachers made out the cards. Such considerations as these would justify reducing the denominator in the formula $\sigma_{D_p} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$ quite considerably, so as to increase the theoretical standard deviation systematically. Another influence which makes for a consistent difference in favor of the observed standard deviations is the inadvertent weighting of various differences of proportions by repetitions of sortings involving practically the same or dependent elements. This is notably the case in the second observed value ($=2.15$). This vitiation crept in before the fact of correlated sortings was noticed.

On the whole, the roughness of these calculations does not hide the very strong and unequivocal support afforded by empirical facts for the theoretical reliability of the percentile method with truly random sampling.

A SERIES OF STANDARDIZED DIAGNOSTIC TESTS FOR THE FUNDAMENTALS OF ELEMENTARY ALGEBRA[1]

Harl R. Douglass
University of Oregon

Standardized tests of progress in the study of mathematics may be of two sorts, those which are primarily tests of power and those which are also tests of rate of work. Mathematical tests and scales, however, have been mostly of the latter type and are given with time limits such that few, if any, pupils are able to do all of the exercises. On these tests the pupil's performance has been described in terms of the number of exercises done correctly and of the number of exercises attempted in a given time. Such tests, with minor exceptions, consist of exercises of approximately equal difficulty, and the tests have been scored on that assumption. For this reason the exercises selected for each test have been of the same type, testing the same ability, and requiring the same degree of power of solution.

The series of standardized diagnostic tests for measuring power in the fundamentals of elementary algebra described in this article are based on somewhat different principles of construction. The intended function of these tests is to measure the power to do certain types of exercises in elementary algebra. In the out-of-school situations in which algebra will be applied, rate of work is clearly secondary to accuracy and within limits negligible. A pupil's performance on these tests is to be described in terms of weighted values of the exercises constituting the test, and within certain large limits to be mentioned later, with no reference to time The exercises of the tests were selected so as to test proficiency in a variety of subtypes of operations under each of the fundamental processes of elementary algebra, and so as to afford considerable range in degree of difficulty. *The tests emphasize a more complete measurement of power and a more thorough and minute opportunity for diagnosis at the expense of measurement for rate of work.*

[1] These tests may be obtained from the University Co-operative Store, University of Oregon, Eugene, Oregon. Price, $1.60 per hundred. Those who use the tests under standard conditions should report median scores and number of cases to the author at Eugene, Oregon.

The fundamentals of algebra.—The first step in the construction of a standardized test for measuring school achievement is the definition of what is to be measured, i.e., in what field of learning or in what portion of a selected field ability is to be measured. In order to answer this question for a test in elementary algebra a questionnaire was prepared that requested those to whom it was addressed to designate the processes of algebra, as ordinarily taught in the first year of secondary schools, which they considered fundamental in the sense that addition, subtraction, multiplication, and division are considered to constitute the fundamental processes of arithmetic. This questionnaire was sent to one hundred members of the American Mathematical Association, whose names were selected so as to include an approximately equal division of teachers in secondary and higher schools, to provide for a representative geographical distribution, and to include those known by the then secretary of the association to be interested in the measurement of mathematical proficiency by standardized instruments.

TABLE I. GEOGRAPHICAL COMPOSITON OF THOSE MAKING REPLIES TO QUESTIONNAIRE TO DETERMINE ELEMENTS OF FIRST-YEAR ALGEBRA

Geographical Location	Number of Teachers		
	Colleges and Universities	Secondary Schools	Total
New England	3	6	9
Atlantic States	6	6	12
Southern States	2	2	4
North Central States	9	12	21
South Central States	2	2	4
Rocky Mountain States	2	2	4
Pacific Coast States	3	2	5
Totals	27	32	59

Fifty-nine replies were received. The geographical distribution of the persons who replied is shown in Table I. More than one-third are from the North Central states but all sections of the country are represented. Table II gives a sum-

mary of the questionnaire. College and university instructors agree very closely with instructors in secondary schools with reference to the operations of first-year algebra which should be considered fundamental. Four processes, (1) collections of terms, (2) division, (3) multiplication, and (4) solution of simple equations received almost a unanimous vote. Four other processes received a majority vote.

. TABLE II. FREQUENCY OF REPLIES TO QUESTIONNAIRE TO DETERMINE FUNDAMENTAL PROCESSES OF FIRST-YEAR ALGEBRA

Fundamental Processes	Number of Teachers		
	Colleges and Universities	Secondary Schools	Total
Collection of terms (Addition and subtraction)	27	32	59
Division	27	31	58
Multiplication	27	31	58
Solution of simple equations	26	28	54
Solution of simultaneous linear equations	19	22	41
Factoring of type forms	19	21	40
Solution of simple quadratics	18	20	38
Graphing	17	17	34
Transposition	13	16	29
Exponential manipulation	11	14	25
Expansion of binomials	12	13	25
Clearing of fractions	11	14	25
Radicals	12	12	24
Ratio and proportion	10	9	19
Evaluation of formulæ	3	3	6
Forming equations	4	1	5

The content of the tests.—For each of the four processes which were most frequently considered fundamental, there was constructed a test consisting of ten exercises. This number was considered an adequate compromise between the desirability of including a large number of exercises for minute measurement and the desirability of a brief test which would not require an undue amount of time for administration. The following considerations in the selection of the exercises were observed:

1. The exercises selected should clearly require proficiency in the fundamental process for which the test was being constructed.

2. The list of exercises in each test should provide for testing the chief subtypes of difficulty and teaching units in each fundamental process.

3. The exercises should be so selected that a differentiation of power would be possible on the basis of the degree of difficulty involved.

4. For the purpose of complete measurement and differentiation each test should contain one or more exercises which could be solved by only a small percent of first-year algebra pupils.

These principles were exemplified in the following tests:

STANDARD DIAGNOSTIC TESTS FOR ELEMENTARY ALGEBRA

Test I. Collection of Terms[1]

1. Add
$$\begin{array}{r} 15\,m \\ 12\,m \\ \hline \end{array}$$

2. Add
$$\begin{array}{r} 8\,ab \\ -7\,ab \\ -5\,ab \\ +2\,ab \\ \hline \end{array}$$

3. Subtract
$$\begin{array}{r} 12\,a \\ 7\,a \\ \hline \end{array}$$

4. Subtract
$$\begin{array}{r} -4\,ab \\ 11\,ab \\ \hline \end{array}$$

5. Add
$$\begin{array}{r} 3a - 8b - 6c \\ 5a + 7b - 3c \\ \hline \end{array}$$

6. Subtract
$$\begin{array}{r} 15a^2 - 8a - 3 \\ 6a^2 + 7a + 2 \\ \hline \end{array}$$

7. Add
   $7a - 6a^2 + 10$ and $a^2 + 6a - 7$
8. Subtract
   $x + xy - 3y$ from $6x - 3xy + 7y$
9. Collect Terms
   $a^2 + 8a + 7a^2 + x + 7 + 9x + 4 + 3a =$
10. Collect Terms
   $2x - xy + 3y - 3x + 2xy + 7y + 8 =$

Test II. Multiplication

1. Multiply
$$\begin{array}{r} 9\,m \\ 2 \end{array}$$

2. Multiply
$$\begin{array}{r} 4\,x^2 \\ 3\,x^2 \end{array}$$

3. Multiply
$$\begin{array}{r} 5\,ab \\ -2\,a^2 \end{array}$$

4. Multiply
$$\begin{array}{r} +3\,ay \\ -\,6a^2y \end{array}$$

5. Multiply
$$\begin{array}{r} 4a - 3b \\ 2\,ab \end{array}$$

6. Multiply
$$\begin{array}{r} 4a^2 - 2x \\ -\,3x^2 \end{array}$$

[1] These tests are printed on a four page folder, 8½ x 11. The exercises are arranged so that sufficient space is provided for the pupil's work.

7. Multiply
$$\begin{array}{r} 8x - 3y \\ 4x + 2y \end{array}$$

8. Multiply
$$\begin{array}{r} 7a^2 - 3x \\ -4ax^2 + 2 \end{array}$$

9. Multiply
$$\begin{array}{r} 3a^2bc + 5b^2cd \\ -4ax^2 + 2 \end{array}$$

10. Multiply
$$\begin{array}{l} 7by^2z - 4a^2cx^2 + 9ab^2y \\ 9byz^2 + 8a^2bc^3 \end{array}$$

Test III. Division

1. Divide $12a^4$ by $2a^3$.
2. Divide $16x^3$ by $-4x^2$.
3. Divide $-18x^2y$ by $2xy$.
4. Divide $-16x^6 + 28x^4 - 24x^3 + 8x$ by $4x$.
5. Divide $20a^3 - 15a^2 - 5a$ by -5.
6. Divide $14x^3 - 28x^2 + 21x$ by $-7x$.
7. Divide $22a^2b^3 - 16a^3b + 8a^2b^4$ by $-2ab$.
8. Divide $36a^3b + 6a^2b^2 - 12ab^3 + 18ab^4$ by $-6ab$.
9. Divide $6a^3 - 18a - 11a^2 + 20$ by $2a - 5$.
10. Divide $r^3 - 19r + 84 - 6r^2$ by $r - 7$.

Test IV. Solution of Simple Equations

1. Solve for x
$$4x = 12.$$

2. Solve for y
$$5y = 20.$$

3. Solve for a
$$4a - 3 = 17.$$

4. Solve for b
$$5b + 6 = 18.$$

5. Solve for r
$$5r - 7 = 63 - 3r.$$

6. Solve for b
$$3b - 6 = 14 - b.$$

7. Solve for y
$$\frac{2y}{3} - \frac{3y}{4} - \frac{y}{6} = 10.$$

8. Solve for a
$$\frac{6a}{7} + \frac{10a}{14} - a = 4.$$

9. Solve for x
$$\frac{2x}{3} + 8 = \frac{3x}{2} + \frac{x}{6} + 14.$$

10. Solve for x
$$\frac{3x}{5} - \frac{x}{2} + \frac{7x}{10} = \frac{1}{2} + 2x.$$

A large number of recent texts in first-year algebra were examined, and exercises were selected therefrom in conformity with the above requirements. For example, in Test II, Multiplication, Exercise 1 is the simplest type of multiplication that may be called algebra, a positive literal by a positive numerical monomial. Exercise 2, somewhat more difficult, consists in multiplying one positive literal monomial by another, involving exponential manipulation, a phase of multiplication. In Exercises 3 and 4, the multiplier is a negative quantity; hence the multiplication involves a new element, that of the laws of signs for multiplication. In Exercise 5 the process is that of finding the product of a positive binomial and a positive monomial; while in

Exercise 6 the multiplier is a negative monomial. In Exercises 7 and 8, the multiplier is also a binomial, Exercise 8 being more difficult than Exercise 7 because of the complexity of the literal factors. Exercises 9 and 10 are more difficult and were inserted simply to differentiate among the abler students.

Method of administration.—Sufficient blank space is provided on the test sheets for all work necessary for the solution of the exercises. A testing of pupils in all of the subtypes of processes and on the varying degrees of difficulty requires that all pupils have an opportunity to attempt all of the exercises. It is evident, however, that for purposes of economy of time pupils may not be permitted to puzzle over the exercises indefinitely. It also seems that the pupils who work extremely slowly do so because of an unsatisfactory grasp on the methods of solution, finding it necessary to retrace some of the work or to spend undue time puzzling over the proper procedure; and they should, consequently, be penalized Hence, a time limit was set on each test which would permit all workers to do all of the exercises provided they knew how to proceed without undue hesitation and did not attempt an exercise more than once. On the basis of these considerations the following time limits were set: Test I, 5 minutes; Test II, 7 minutes; Test III, 9 minutes; Test IV, 8 minutes

Since the purpose of the tests is not to ascertain how many exercises may be done in a given time, and since the emphasis is placed upon accuracy rather than speed, a measure of proficiency may be secured which is not affected by the element of rate of work, and which more closely resembles the demands of the life situations in which algebraic manipulations in question will be applied. A psychological element of "hurry" is not aroused, and the likelihood of pupils failing to solve exercises because of excitement or undue speed is reduced materially.

The directions to the pupils suggest that they are to work carefully rather than hurriedly. They are told: "We want to see how many of these exercises you can do correctly. You will be given time enough to try all the exercises if you do not waste any time. Work rapidly but do not hurry. Accuracy will count more than rapid work."

Weighting the exercises.—The tests were given to 938 pupils in fourteen high schools in five states of the middle west and far west. The distribution of scores, the medians, 25- and 75-per-

TABLE III. DISTRIBUTION OF PUPILS ACCORDING TO THE NUMBER OF EXERCISES SOLVED CORRECTLY

Number Exercises Solved	Tests			
	I	II	III	IV
0	10	1	4	16
1	12	5	53	13
2	34	10	53	38
3	70	17	62	47
4	56	19	64	104
5	118	74	71	142
6	144	110	103	238
7	152	144	150	139
8	130	241	138	85
9	128	217	99	59
10	84	50	101	57
Total number tested	938	888	898	938
25-percentile	4.44	5.87	3.82	5.10
Median	7.16	8.27	7.26	6.46
75-percentile	8.88	9.21	8.82	7.76
Quartile deviation	2.22	1.67	2.50	1.33

centiles and quartile deviations, from these papers scored without weighting the exercises of the tests, are shown in Table III. Weights for the exercises of the tests were determined by reference to the curve of probability and expressed in terms of median deviation (P.E. or M.D.). A zero point for each test was determined by methods in recognized use.[3] The weights with reference to these zero points are given in Table IV.

Methods of scoring.—In out-of-school situations where algebraic processes are required, the test of proficiency is accuracy. Answers are either right or wrong. No leniency is shown in business dealings to the individual who "has the right method but who has made a small mistake." For this reason and to facilitate scoring, solutions are graded either as right or as wrong. Full credit or no credit is allowed. Answers being correct except for arrangement of terms are not counted wrong. The sum of the weights of the exercises done correctly constitutes the score of the pupil.

[3] Hotz, H. S. *First-year algebra scales.* (Teachers College Contributions to Education, No. 90, 1918), pp. 62–70; also Trabue, M. R. *Completion test language scales.* (Teachers College Contributions to Education No. 77, 1916), pp. 45–60.

TABLE IV. FINAL WEIGHTS IN P. E. OF EACH EXERCISE OF EACH TEST. (TO NEAREST TENTH ONLY)

Exercise	Tests I	Tests II	Tests III	Tests IV
1	1.4	1.7	1.1	2.0
2	1.6	2.7	2.3	1.8
3	1.8	2.9	2.7	3.1
4	3.3	3.2	2.6	4.0
5	2.0	3.6	2.9	4.5
6	2.8	3.0	2.9	3.6
7	2.6	3.9	3.3	5.9
8	3.1	4.6	3.2	5.7
9	3.9	5.1	3.5	6.0
10	3.5	6.4	3.4	6.0
Total	26.0	37.1	27.9	42.6

Standard scores.—Standard median scores for late in February or early in March taken from 938 papers written by first-year classes in fourteen high schools in five states of the West and Middle West, are as follows: Test I, 16.2; Test II, 24.0; Test III, 16.9; Test IV, 20.0.

Values and limitations of the tests.—These tests possess both certain values and certain limitations; chief among which are:

VALUES:

1. The tests provide adequate opportunity for diagnosis of weakness.
2. Opportunity is provided the classroom teacher to check the effectiveness of teaching with respect to the various types of the four fundamental processes.
4. Accurate measurement of power is possible because of the variety in types of exercises and degrees of difficulty.

LIMITATIONS:

1. But one or two exercises are included for each subtype of operation.
2. Because of the extended time limits, rate of work cannot be adequately measured.
3. The tests apply only to the four fundamentals of elementary algebra as determined by the combined judgments of those replying to the questionnaire.

SCALE OF ATTAINMENT NO. 3.—FOR MEASURING "ESSENTIAL ACHIEVEMENT" IN THE THIRD GRADE

LUELLA W. PRESSEY
Ohio State University

ANALYSIS OF THE CURRICULUM OF THE THIRD GRADE

Reports on two previous "Scales of Attainment" have already appeared.[1] The first of these was an examination for the second grade, the second for the eighth grade. The idea back of all three of these examinations is essentially the same:—the purpose of measuring the *essential achievement* of the grade. Only such subjects as could properly be called "promotion" subjects have been considered, since others, though important from a cultural standpoint, hardly condition a child's school progress. The curriculum of the third grade includes spelling, reading, arithmetic, drawing, singing, and writing. Of these drawing and singing are of little importance from the writer's standpoint since ability in these subjects does not affect the progress of a child through the grades. Also, children are not generally retarded because they cannot write well. The fundamental "promotion" subjects of the grade appear then to be spelling, reading, and arithmetic. And the question, in constructing this examination for the third grade, was as to the most desirable form for tests in these three subjects.

Words for the spelling test could, of course, be obtained from the Ayres' scale. The troublesome problem was whether to have list spelling, a timed sentence spelling test, or some other special form. The construction of a test in silent reading was a more difficult matter. Under the direction of the writer, a special study was made by the members of a large extension class. It was concluded that an ability to read with sufficient understanding to grasp story values was the most important factor for this test. Children in the third grade should be masters of the technic of reading to such a degree as to be able to pay attention to the ideas presented, provided the material is reasonably simple. They

[1] Pressey, L. W. "Scale of attainment No. 1.—An examination of achievement in the second grade," *Journal of Educational Research*, 2: 572-81, September, 1920; Pressey, S. L. "Scale of attainment No. 2.—An examination for measurement in history, arithmetic, and English in the eighth grade," *Journal of Educational Research*, 3: 359-69, May 1921, complete.

should be able to read short stories with understanding and with a certain grasp of the story as a whole. The point most emphasized was that the reading matter selected for the test should be coherent. Isolated sentences, bits of incidents, short selections without any particular story value, would not give measures of greatest use to the teacher, since they would not demand of the child that he grasp the story.

Measurement in arithmetic remained to be considered. In the third grade there is intensive drill in the fundamental operations. But there is more than this; a real beginning is made in solving problems. The children should learn to read a simple problem, interpret it for themselves, and apply the needed process. So it appeared that both skill in the fundamental operations and ability to solve simple problems should be involved in any test of third-grade work in arithmetic. Though the proportion of problem work to drill in fundamentals varies a great deal from one teacher or from one school to another, all teachers and all schools seem to have some work in both fields.

It was decided that the test in spelling should involve the spelling of words in their proper setting in a sentence, that the reading test should require the grasping of the meaning of a coherent story, and that the arithmetic test should involve both the fundamental operations and simple problems. Upon the basis of this analysis the scale was built.

The Scale and Its Construction

The scale appears on a four page folder, each page 9 by 6 inches. The spelling test—with the lines for name, age, etc.—is on the first page. It is so placed because the children cannot know what to do on this test until told by the teacher and because study of the page in advance by those first receiving their papers is therefore of no advantage. The arithmetic test is on the back page and is given immediately after the spelling. The reading test occupies both of the middle sheets, the items being so arranged that the one at the end of the first page is continued on the second, thus starting the children on the second sheet almost without their knowing it.

Directions for the test appear on a single sheet of the same size. All the directions, except those for the spelling test, are of the "question-and-answer" type, since this method has proved

more successful with young children than anything more formal.[2] The spelling test requires about five minutes to give; the arithmetic test, seven minutes; and the reading test, eight minutes. The children thus work twenty minutes; and the time needed for giving directions, passing out blanks, etc., makes up a total of about half an hour for the entire examination.

Scoring of the tests is reasonably simple. The spelling test takes only about half a minute to rate; the arithmetic test requires about the same amount of time, and the reading test no longer. In all, it is possible to rate the test and obtain the total score in about two minutes. A teacher can thus rate the blanks for a class of thirty-five in not much more than an hour. The total score on the examination consists simply of the sum of the scores on the three tests.

The Spelling Test

In regard to form, the spelling test is somewhat unusual. It consists of sentences such as the following:

1. My ——— told me to go.
2. It is very cold in ———.
3. She has a great deal of ———.

The teacher explains to the children that each sentence on the page has a word left out, and that she will tell them what word they are to write in. She then reads the sentence as it should be, tells which word is to be written, and then repeats the whole sentence. Thus, the directions for the three sentences given above are:

Look at the first sentence. It should be "My mother told me to go." "Mother" is the word you should write on the line where it has been left out. "My mother told me to go."

The next sentence should be "It is very cold in winter." "Winter" is the word you should write. "It is very cold in winter."

The next sentence should be "She has a great deal of money." "Money" is the word you should write. "She has a great deal of money."

By giving the word in its context, the child's attention is not entirely centered upon the spelling of the word. Nothing has been said about a spelling test, nor is there anything on the page to indicate to the children that they are being tested specifically in spelling; they are simply to "write in the word that is left out."

[2] Pressey, L. W. "A group scale of intelligence for use in the first three grades," *Journal of Educational Psychology*, 10: 297-308, September, 1919.

This setting of the spelling problem also gives the children the word in its context without making them write an entire sentence. The method would seem to combine certain of the merits of both list and sentence spelling, since the writing of only one word is required, yet the words are given unmistakably in their context. It should be added that the words were selected from the Ayres list, Columns G to Q.[3]

The Arithmetic Test

The arithmetic test represents an attempt to eliminate writing in the solving of problems since it is ability in arithmetic, not speed of writing, that is being measured.[4] The scheme has been adopted of asking a question, followed by four answers. Only one answer is correct, and the child is to make a mark around it. Sample problems from this test are given below:

6. How much is 8 − 3? **6 5 11 24.**
10. How much is 6 x 3? **18 12 9 3.**
13. How much is 13 + 0? **12 11 14 13.**
16. How much is 33 + 6? **9 39 12 33.**
20. If you bought a bottle of ink for 10c and some candy for 4c, how much would you spend in all? **10c 6c 14c 40c.**
24. Doris has 7 books, Frank has 3 times as many. How many has Frank? **14 21 10 4.**

As is evident from the type of question it is expected that the children will do most of the work in their heads. In fact, on the papers of over a hundred children recently tested there were only two or three marks in the margins. The first four questions are used as examples; two of these deal with the fundamental combinations and two with simple problems. The children seem to

[3] The sentences are not arranged in the order of difficulty of the words to be spelled. The reason for this is psychological. If the list is in order of difficulty, the child feels that the words are constantly becoming harder. He senses that the worst is yet to come. Glancing down his paper, he finds that there are still 10 more lines and he knows that if the words keep on getting harder he will shortly be beyond his depth. As a result he becomes discouraged and misses some of the words he might really spell if he were not so apprehensive. It is surely of no particular importance—since the entire list is given in any case—that the words should appear in any special order, provided the *same* order be used for all.

[4] Thorndike, E. L. and Courtis, S. A. "Correction formulæ for addition tests," *Teachers College Record*, 21: 1-24, January, 1920.

find no difficulty with this method of presenting the problems.[5] In fact, they seem able to work without fatigue for a considerably longer time than is possible by the usual method of working out the problems on paper. The saving in time is also of advantage.[6]

It seems reasonable, then, to suppose that the form of the test is satisfactory. It is very certain that the children like it, it does not have a bad psychological effect upon them, as is the case with other arithmetic tests with which the writer is acquainted, and it appears to be completely applicable to this grade.

The Reading Test

The reading test is more nearly similar to other standardized forms than is the case with the spelling and arithmetic tests. It has, however, from a psychological point of view, one rather important difference. It is made up of paragraphs each of which relates a fairly coherent story or incident. Following each paragraph are four questions. The questions are answered in just the same way as with the arithmetic problems; that is, the child selects the correct answer—from four possible ones—by drawing a line around it. Being already familiar with this type of response as exemplified in the test just preceding, the children find no difficulty in understanding what they are to do. A sample passage appears below:

Once a bright star wanted to come down to earth and be a flower so that she might be near little children who would love her. First, she tried being a white rose, but the children were afraid of the thorns. Then she tried being a daisy, but the children didn't see her because she was so small. Then she

[5] It will be noticed that the wrong numbers are those which will be obtained if the wrong process is used. Thus, in the 20th item presented above, if the child subtracts or multiplies (he can't divide) instead of adding, he will obtain either 6 or 40 as an answer. Both these numbers appear in the list, so he will not be deterred from his wrong idea for lack of an answer that agrees with it. As far as possible, wrong answers have been anticipated in this way.

[6] In an arithmetic problem test recently devised there are 10 problems; the children (of the third grade) are allowed to work for a half hour on these ten. Certainly such a test takes on the character of a measure of endurance. It also leaves out, to a very great extent, the element of speed which is of importance in doing arithmetic. Readiness in grasping a problem, quickness in making the necessary combinations, are surely as essential to real ability in arithmetic as the power to solve problems provided one is given time,—for in real life (for which the school is presumably preparing) one certainly has to make change and perform other simple mathematical operations within a very short time limit.

went up on the cliffs and became a dew-drop, but the children could not climb so high. Finally, she lighted softly on the waters of a shallow pond and became a beautiful white water-lily. Then the star was happy because she was near little children and they loved her.

1. What did the star like best to be? **rose dew-drop water-lily daisy.**
2. What did the star want to be near? **cliffs pond children earth.**
3. What flower were the children afraid of? **rose daisy pansy lily.**
4. Where did the dew-drop grow? **pond cliff field lake.**

This passage should be compared with similar passages from other reading tests for this grade to appreciate the difference in "story value" between it and those appearing in other tests. The selections given below are from various reading tests that employ a more or less similar method.

A crab who lived in a sand-hill was sitting at his door in the sun eating a rice cake. An ape went by carrying an orange seed.
Where did the crab live? ....

John had two brothers who were both tall. Their names were Will and Fred. John's sister, who was short, was named Mary. John liked Fred better than either of the others. All of these children except Will had red hair. He had brown hair.

1. Was John's sister tall or short? ....
2. How many brothers had John? ....
3. What was his sister's name? ....

This book is lying on the desk (a picture of a book, face open, is presented just above the paragraph), but it is hard to make it stay open. With your pencil draw a single straight line to represent a ruler lying across the book to hold the pages open. Be sure to make the line from one side to the other, across the book, instead of making it go up and down.

It is the writer's contention that the above passages have no story value. They are simply isolated *reading exercises*. If the reading matter of a passage doesn't tell a story or develop an idea, what is the use of being able to read it? Such reading matter exists—fortunately—for the most part, only in tests. In all other reading there is usually a story, a description, or an exposition, of some sort of significance. Again, the question of gaining the interest of the children is a vital one. Surely no child could be really interested in such selections as those given above. They don't start anywhere, and they don't arrive anywhere, and they aren't about anything; indeed, it would take a statistician to identify correctly the family described in the second passage. There is a distinct movement now on foot to give children in school reading matter that there appears to be some interest in

reading; such tests as the above are certainly not in the spirit of this movement. Incidentally, there is considerable doubt as to whether the ability to gain a coherent idea of the passage read is at all tested by these tests. They would seem rather to test a type of mental alertness.

Results and First Norms

The separate tests of the scale were first tried out by teachers in university extension classes. From these results, items were selected and a trial form of the scale printed. This trial form was given as part of a survey of Bedford, Indiana.[7] The results showed the need of certain revisions which were accordingly made in the final form which is now ready for distribution.[8] Since making the final form the following tentative norms have been obtained (October testing):

Norms (October)

	Median	No. Cases
Spelling	12	198
Arithmetic	9	198
Reading	11	198
Total Score	30	198

The writer has done very little in the way of validation of the scale. But the test most in need of validation has perhaps been sufficiently investigated. Several teachers have been asked to make ratings of their pupils as to their ability to read with understanding. Correlations of the reading test results with these ratings vary from +0.60 to +0.82. This would seem to afford considerable evidence that the test is really measuring the ability of the children to comprehend what they read.

About the spelling test, there can be little question. Children in the third grade do little in the way of using spelling as a tool in writing. Accordingly, sentence spelling would hardly be applicable to the grade. The present test is presumably at least as good

[7] The writer wishes to express her indebtedness to Superintendent E. W. Montgomery of Bedford, Indiana, for his cooperation in this and many other problems.

[8] The blanks may be obtained in quantities from the Department of Psychology of Indiana University at $0.90 per 100.

a measure of spelling ability as list spelling—and it avoids some of the objectionable features of the latter. There might be some question about the arithmetic test. There is surely little difference between the material used in the test and that used in the school room; any difficulty would be with the method of indicating the answers. Since no child has yet made a zero score on this test, the method of indication would not seem to be so difficult as to interfere seriously with the solution of the problem.

Certain Fundamental Factors Concerned with Test Building

In conclusion, the writer would like to point out one or two rather fundamental aspects of testing and of test construction in the elementary grades which are exemplified in this scale, but which should appear in all scales for use in these early grades—and perhaps in the upper grades as well.

1. One of these fundamental points is the matter of motivation. The writer feels sure that motivation by interest is the only reasonable, indeed the only possible, way with young children. A test that relies upon school discipline for its motivation cannot but be unsound from a psychological standpoint. A motivation coming from interest vitalizes the test situation as nothing else can, and should be sought after wherever possible.

2. The participation of the teachers themselves in the making of a test, and in the selection of the materials for it, is of great importance. They are intimately in touch with the teaching situation. The test builder sometimes, to judge from his tests, is not. There are at present many scales on the market that are rightly condemned by teachers, because they are not closely in touch with the teaching, or are not adapted to the children, or are not fitted for use by the teachers. In the present instance, the writer has contributed little save the method of presenting the material and the technical part of test construction. The rest is the result of the study and observation of competent teachers in test making. A much greater participation of those who are going to use the tests is to be hoped for.

3. Achievement tests in "battery" form are now becoming popular. In such an instrument, tests in more than one subject are included. Under such circumstances, it is easy for an examination to become elaborate and cumbersome. The manuals of directions

for using the test may become prolonged to thirty or forty pages; the time required to make preparation for the giving and the scoring of the test amounts to hours; and the interpretation of the scores finally obtained becomes so involved as to need a personally conducted tour through the results. It has been the writer's intention to keep clear of such elaborateness in giving, scoring, handling, and interpreting. It is possible to keep such achievement scales simple and easy to use. He who does not do so is giving an example of intellectual indolence and poor workmanship. Elaborate scales can be avoided; but an earnest teacher is likely to be led into the use of them if she does not first stop to reckon up the amount of time she must invest in the performance.

Summary

The paper presents a brief scale of attainment for measuring the essential progress in the third grade.

1. Tests in reading, spelling, and arithmetic are included. The form of these tests is, in some respects, new.

2. Validation and first norms are presented.

3. Three suggestions are made concerning fundamental aspects of testing, especially in the early grades: (a) that the test motivation come from interest rather than school discipline, (b) that teachers should be allowed to participate in the building of tests, and (c) that achievement scales should, and can, be kept sufficiently simple in construction to be of great use to teachers.

AN ANALYSIS OF THE CONTENT OF SIX THIRD-GRADE ARITHMETICS

F. T. Spaulding
Scarborough School, Scarborough, New York

Recent investigations of the content of the elementary-school course of study have centered attention on the utility and effectiveness of the subject-matter which the course of study includes. The Committee on Minimum Essentials and the Committee on Economy of Time in Education, as well as numerous individual investigators, have proposed more or less definite standards for each of the elementary-school subjects. The present study of six third-grade arithmetics was intended to determine the extent to which certain well-known primary texts approach these standards.

Specifically stated, the purpose of this study was twofold: first, to determine the exact nature of the arithmetical work presented; and second, to provide a basis for a judgment of the extent to which the textbooks studied make an appeal (a) to the immediate needs and interests and (b) to the probable future needs and interests of the pupils using them.

In selecting the textbooks for study an effort was made to choose those which are in wide use at the present time or which have been widely used in the past. The six books selected represent a period of fourteen years in the development of arithmetic texts. Listed according to recency of publication they are:

Stone, J. C., and Millis, J. F.: *New Stone-Millis Arithmetic.* (Primary.) Benj. H. Sanborn & Co., 1920.

Chadsey, C. E., and Smith, J. H.: *Efficiency Arithmetic.* (Primary.) Atkinson, Mentzer & Co., 1917.

Hoyt, F. S., and Peet, H. E.: *Everyday Arithmetic.* (Book I.) Houghton Mifflin Co., 1915.

Walsh, J. H., and Suzzallo, H.: *Walsh-Suzzallo Arithmetics.* (Third Year.) D. C. Heath & Co., 1914.

Wentworth, G., and Smith, D. E.: *Arithmetic.* (Book I.) Ginn & Co., 1911.

Milne, W. J.: *Progressive Arithmetic.* (First Book.) American Book Co., 1906.

The method pursued was substantially that used by Wise[1] and Monroe[2] in their studies of arithmetic problems. The third-grade material presented in each book was divided into two classes—(1) examples, i.e., drill and test work, in which the operations to be performed were indicated for the pupil; and (2) problems, or work in which the method of solution was not directly indicated. To make the results as nearly comparable as possible, each separate example or problem was counted as a unit, even though the author had included several such in a single task. Tables of numbers, from which the teacher or the pupil could form an almost limitless series of examples, were omitted in the classification.

Examples and problems were then classified separately as to the arithmetical operations or combinations of operations involved in each, including the use of fractions. Where fractional numbers were employed merely to indicate division (as ½, ¼, etc.) the examples and problems were classified under "Division," classification as "Fractions" being restricted to examples and problems using fractions as such.

Problems were further classified (1) as to general subject-matter, according to a modification of the scheme adopted by Monroe, and (2) as to the use of measurements and the types of measurements employed.

The totals of examples and problems found in the six books, with the proportions of each, are presented in Table I. It is noteworthy that the totals vary from 1777 (Chadsey-Smith) to 3106 (Wentworth-Smith)—a range of nearly 1400. There has been little standardization, apparently, of the amount of material to be covered in this grade. Wide variation is also evident in the proportions of examples and problems presented. There would seem to be a tendency, however, toward a greater proportion of problems in the more recent books—a hopeful sign (assuming that the problems are of the right type), in view of the present emphasis on the need for making schoolwork vital and concrete, rather than abstract.

[1] Wise, C. T., "A survey of arithmetical problems arising in various occupations." *Elementary School Journal*, 20: 118-36, October, 1919.

[2] Monroe, W. S., "A preliminary report of an investigation of the economy of time in arithmetic." *Sixteenth Yearbook of the National Society for the Study of Education, Part I*, chapter 7.

TABLE I. NUMBERS AND PROPORTIONS OF PROBLEMS AND EXAMPLES CONTAINED IN SIX THIRD-GRADE ARITHMETICS

Texts	Number			Percent	
	Examples	Problems	Total	Examples	Problems
Stone-Millis (1920)	1,526	751	2277	67.0	33.0
Chadsey-Smith (1917)	1,334	443	1777	75.1	24.9
Hoyt & Peet (1915)	1,887	667	2554	73.9	26.1
Walsh-Suzzallo (1914)	2,513	582	3095	81.2	18.8
Wentworth-Smith (1911)	2,344	762	3106	75.5	24.5
Milne (1906)	2,059	442	2501	82.3	17.7
Average	1,944	608	2552	76.2	23.8

Table II shows the distribution of examples according to the types of operations involved. With the exception of Milne, the books are practically agreed in requiring the use of a single one of the four fundamental operations in the solution of over 90 percent of the examples and of more than 85 percent of the problems—a practice supported by the findings of Wise in his study of problems taken from everyday experience. But in the relative amount of space devoted to each of the four single operations the books vary widely—so widely that the median is of little value as an indication of general practice. Here again practice has apparently been determined in each case by the arbitrary judgment of the textbook maker, rather than by reference to any sort of objective standard.

The variation in proportionate emphasis on the fundamentals between examples and problems within the same book is as worthy

TABLE II. PERCENTS OF EXAMPLES AND PROBLEMS INVOLVING THE VARIOUS OPERATIONS

Combinations	Stone-Millis			Chadsey-Smith			Hoyt & Peet			Walsh-Suzzallo			Wentworth-Smith			Milne			Medians		
	Examples	Problems	Total	Examples	Problems	Total	Examples	Problems	Total	Examples	Problems	Total	Examples	Problems	Total	Examples	Problems	Total	Examples	Problems	Total
Addition	26.7	20.5	24.7	22.0	26.5	23.1	44.4	21.3	38.4	22.3	21.1	22.1	32.5	19.7	29.3	12.0	5.4	10.8	24.5	20.8	23.9
Subtraction	15.1	12.9	14.4	13.0	10.8	12.5	29.3	19.8	26.7	13.5	17.7	14.3	27.3	21.5	25.9	10.8	7.2	10.2	14.3	15.3	14.4
Multiplication	28.2	36.6	31.0	25.4	18.5	23.6	10.7	21.4	13.5	29.4	32.0	29.9	13.2	30.6	17.6	21.9	27.0	22.8	23.7	28.8	23.2
Division	28.4	27.6	28.1	36.3	32.5	35.3	15.6	23.8	17.7	29.7	22.7	28.4	18.8	20.5	19.2	33.0	12.7	29.4	29.1	23.3	28.3
Addition and Subtraction	....	0.3	0.1	....	0.5	0.1	....	5.5	1.5	0.1	....	....	....	0.4	0.1	0.4	2.7	0.8	0	0.5	0.1
Addition and Multiplication	1.5	0.1	1.1	....	1.1	0.3	....	4.0	1.0	....	0.5	0.1	5.5	2.5	4.8	3.1	8.1	4.0	0.8	1.8	1.1
Addition and Division	....	0.1	....	....	....	....	....	1.4	0.4	....	0.7	0.1	1.8	0.2	1.4	....	....	....	0	0.4	0.1
Addition, Subtraction & Multiplication	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	.2	....	0	0	0
Addition, Subtraction and Division	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Addition, Multiplication and Division	....	0.1	....	....	1.8	0.5	....	....	....	....	....	....	....	0.1	....	....	....	....	0	0.1	0
Subtraction and Multiplication	....	0.3	0.1	....	....	....	....	2.1	0.6	....	0.7	0.1	....	....	....	0.7	7.0	1.8	0	0.5	0.1
Subtraction and Division	0.1	....	....	....	....	....	....	0.5	0.2	....	0.3	0.1	....	....	....	....	....	....	0	0	0
Subtraction, Multiplication and Division	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Multiplication and Division	....	1.5	0.5	....	0.5	0.1	....	0.2	....	....	0.2	....	....	....	....	....	11.1	2.0	0	0.4	0.1
All Operations	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Fractions	....	....	....	3.3	7.9	4.5	....	....	....	5.0	4.1	4.9	0.9	4.5	1.7	18.1	18.6	18.2	2.1	4.3	3.1

of note as the variation between books. Leaving out of consideration a fairly uniform preponderance of addition examples over addition problems (due partly to the use of addition in demonstrating subtraction and multiplication), the tables show in some cases a greater proportion, in others a less, of examples than of problems. Hoyt & Peet, for instance, devote 29 percent of their examples and only 19 percent of their problems to subtraction; whereas 8 percent more problems than examples are devoted to division. In the *Walsh-Suzzallo Arithmetics* the differences are in the opposite direction. Part of the discrepancy is obviously due to the fact that considerably more problems than examples are distributed in the double and triple-operation groups; but this would provide only for smaller proportions of problems than of examples. Lack of a clear conception of the functions of problems and of examples and of the amount of practice desirable in each is responsible for most of the variation.

A hopeful sign is apparent in the partial or the complete elimination of fractions in the later books. Milne devotes 18.2 percent of his examples and problems to fractions as such; Stone & Millis and Hoyt & Peet postpone the study of fractions (except fractions with unit-numerators, used to indicate simple division) to a later grade. In his study of common problems from everyday life, Wise found the use of fractions to be restricted very largely to that of fractions with simple numerators and denominators—a fact which would argue for much less attention to the more complex fractions than has been given by the older arithmetics.

From the point of view of the textbook maker such investigations as that of Wise must serve to indicate the nature of the material which should be included in a course of study, rather than the relative stress which each phase of such material should receive. For it is obvious that the practice necessary to master an operation of arithmetic may be out of proportion, so far as the time element is concerned, to the degree to which the pupil may expect to use this operation in ordinary life; and yet the fact that it is necessary in ordinary life requires its mastery in school. It is dangerous, therefore, to draw conclusions from such figures as we have at hand as to the exact amount of time which should be devoted to each of the several phases of arithmetic-teaching. In so far as the textbooks studied center their attention on practice in the fundamentals they are in accord with the general

principles established by Wise's study. Their weakness is in pedagogical method, rather than in aim; they show no evidence of established standards as to the amount of practice needed to master the fundamentals.

In the classification of the subject-matter of the problems (Tables III, IV, and V), we find a basis for determining the extent to which the texts meet the present and the future needs and interests of the pupils who use them. Here again a very wide diversity is apparent. Of particular significance is the variation in the proportion of problems which relate to no human activity—such problems as, "How many feet have 16 birds?" or "A book cost \$1¾. How much more than \$1¼ is that?" The Hoyt & Peet *Everyday Arithmetic* is the only one which has a clean slate in this respect; the Walsh-Suzzallo text is the worst offender, with 4.1 percent of its problems (24 in all) of this worthless type. The tabulation shows an encouraging tendency, however, toward the elimination of this sort of work from the later books.

There is a much larger group of problems (49.6 percent of those in the Wentworth-Smith book) which give valuable practice to the pupil and are of a sort which he will frequently meet, but which as presented can be identified with no particular activity. Of this type are such problems as, "How many pecks are there in 8 bushels?" or, "What is the perimeter of a lot 180 feet square?" In their lack of appeal to the pupil and their lack of connection with concrete activity, they are of much less value, of course, than problems which possess definite significance. As with the problems relating to no activity, the proportion of such problems seems to be decreasing in the more recently published books.

With the exception of Wentworth-Smith the books are fairly agreed in devoting about half their problems to Home Activities, Personal Activities, and the Activities of Children. Five books out of the six agree in relating about 10 percent of their problems to Home Activities; in the other two classifications just mentioned there is considerable divergence. Within all of these fields the quality of the problems presented counts for so much that without a more exact classification than that of the present study, and without more definite standards of evaluation than have yet been developed, we can draw no worth-while conclusions as to present tendencies and limitations.

For the distribution of problems dealing with occupations we have standards of a sort. If we consider these problems in the

TABLE III. DISTRIBUTION OF PROBLEMS ACCORDING TO SUBJECT-MATTER

Texts	1. Occupations														2. Home Activities		3. Personal Activities		4. Activities of Children		5. Relating to No Activity		6. Incapable of Identification	
	1. All Occupations		A. Agriculture		B. Trade		C. Transportation		D. Public Service		E. Industry		F. Others											
	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent
Stone-Millis	240	31.8	97	12.9	113	15.0	8	1.1	10	1.3	1	0.1	11	1.4	83	11.1	56	7.5	249	33.2	5	0.7	118	15.7
Chadsey-Smith	59	13.3	18	4.1	10	2.3	13	2.9	8	1.8	1	0.2	9	2.0	40	9.0	57	12.9	185	41.8	1	0.2	101	22.8
Hoyt & Peet	195	29.5	41	6.1	81	12.2	29	4.4	27	4.2	10	1.5	7	1.1	75	11.2	151	22.6	191	28.5	0	0	55	8.2
Walsh-Suzzallo	140	24.1	61	10.5	62	10.7	10	1.7	1	0.2	6	1.0	0	0	72	12.4	103	17.7	141	24.2	24	4.1	102	17.5
Wentworth-Smith	74	9.6	17	2.2	46	6.0	4	0.5	4	0.5	0	0	3	0.4	82	10.8	125	16.4	79	10.4	24	3.2	378	49.6
Milne	89	20.2	11	2.5	50	11.3	10	2.3	4	0.9	0	0	14	3.2	77	17.4	51	11.5	80	18.1	7	1.6	138	31.2
Average	132	21.4	41	6.4	60	9.6	12	2.2	9	1.5	3	0.5	7	1.2	72	12.0	91	14.8	154	26.0	10	1.6	149	24.2

light not merely of the pupils' present needs, but of their future interests and of the possibility of vocational and civic enlightenment through the work in arithmetic, we find not only much variation in individual texts, but a considerable difference in treatment between the earlier and the later books. The subject-matter in the problems in the earlier books appears to have been selected almost entirely because of its adaptability to the text-book maker's purposes in providing practice in the fundamentals, rather than because of any consideration of the value of the subject-matter itself. The grouping of problems is for the most part heterogeneous and pointless, so far as their subject-matter is concerned. In the later arithmetics it has been the avowed purpose of the authors, attested not merely by their introductory statements but by the arrangement and selection of problems as well, to provide subject-matter which shall make definite appeal to the needs and interests of the pupils. In these later books we find, therefore, first, a distribution of problems over a wider field of interests and activities, and second, a more thoughtful apportionment of problems to the various occupational groups. The first characteristic becomes apparent when we consider the percents of problems represented in each group. The three last published books afford problems under all six occupational headings, whereas the earlier texts tend seriously to neglect various fields—

TABLE IV. A COMPARISON OF THE DISTRIBUTION OF PROBLEMS ACCORDING TO OCCUPATIONS, WITH THE DISTRIBUTION OF THE WORKING POPULATION OF THE UNITED STATES

	All Occupations	A. Agriculture	B. Trade	C. Transportation	D. Public Service	E. Industry	F. Others
Census of 1910	*100.0*	*33.2*	*9.5*	*6.9*	*1.2*	*27.9*	*21.3*
Stone-Millis	100.0	40.4	47.1	3.3	4.2	0.4	4.6
Chadsey-Smith	100.0	30.5	17.0	22.0	13.6	1.7	15.2
Hoyt & Peet	100.0	21.0	41.5	14.9	13.9	5.1	3.6
Walsh-Suzzallo	100.0	43.6	44.3	7.1	0.7	4.3	
Wentworth-Smith	100.0	23.0	62.2	5.4	5.4		4.0
Milne	100.0	12.4	56.2	11.2	4.5		15.7

TABLE V. DISTRIBUTION OF PROBLEMS INVOLVING MEASUREMENTS

Texts	Problems Dealing with Measurement		Linear		Square		Cubic		Land		Dry		Liquid		Weight		Time	
	No.	Per-cent of all Prob.	No.	Per-cent[a]	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent	No.	Per-cent
Stone-Millis	214	29	24	11.3	8	3.4	..	....	..	....	41	19.2	65	30.4	36	16.9	40	18.8
Chadsey-Smith	83	19	26	31.3	..	....	..	....	..	....	15	18.1	13	15.7	7	8.4	22	26.5
Hoyt & Peet	154	23	34	22.1	..	....	..	....	..	....	.	....	12	7.8	9	5.8	99	64.3
Walsh-Suzzallo	139	24	29	20.8	2	1.5	..	....	..	....	18	13.0	20	14.4	26	18.7	44	31.6
Wentworth-Smith	168	22	30	17.9	28	16.7	..	....	..	....	27	16.1	24	14.3	22	13.0	37	22.0
Milne	116	26.2	20	17.2	9	7.8	22	19.0	..	....	..	....	15	12.9	9	7.8	41	35.3
Median	147	24	28	19.4	5	2.5	0	0	..	....	17	14.6	18	14.4	16	10.7	41	29.1

[a] Percent of measurement problems.

notably Agriculture and Industry (Wentworth-Smith and Milne), Transportation (Wentworth-Smith), and Public Service (Walsh-Suzzallo and Wentworth-Smith). The second point is confirmed by an inspection of Table IV, in which the percents of occupational problems belonging to each group are presented for comparison with the distribution of workers by vocations according to the census of 1910 (adapted from Monroe). Though the distribution of problems diverges to a considerable extent from that of the industrial population, there is yet evident in the later books some degree of approach to this standard distribution. The over-emphasis on trade is not a serious defect, since this is an occupation with which every adult (and nearly every child) has more or less to do. The most serious neglect appears in the field of industry, which, though claiming nearly 28 percent of the nation's workers, is represented at most by but 5 percent of the problems.

Our final analysis is concerned with the number of measurement problems and with the types of measurement involved (Table V). With but one exception (Chadsey-Smith), the six textbooks are fairly agreed in devoting about one-fourth of their problems to some type of measurement. Beyond this point agreement ceases. There is a tendency in the later books to postpone study of certain forms of measurement (notably square and cubic measure), taken up in detail in the earlier published texts; but in the amount of space devoted to the remaining forms of measurement (linear, dry, liquid, weight, and time) the books are very widely at variance. There is here very evident need of standardization of the course of study on the basis both of pedagogical and of utilitarian values.

Summary

The very limited number of texts involved in the present study makes the drawing of general conclusions a dangerous proceeding. Until such time as a more extensive investigation is possible, however, the following conclusions are advanced for what they are worth.

1. There is much need for standardization of third-grade arithmetic texts: (a) as to the nature and amount of material presented for study in the course of the year; (b) as to the emphasis placed on each of the fundamental operations; and (c) as to the subject-matter of the problems presented for solution.

2. Present practice in textbook writing apparently tends toward a concentration of attention in the third grade on the fundamental operations of arithmetic, with a postponement of the study of fractions to the upper grades—a tendency the value of which is supported by the findings of Wise and of others with respect to the utility of these phases of arithmetic in practical experience.

3. There is evident also a tendency toward greater emphasis on problem-solving, in contrast to the simple "doing examples" of the earlier books.

4. A study of the subject-matter of problems shows an increasing elimination of those which relate to no human activity, or which can be identified with no activity—and, as a corollary, an increase in the proportion of worth-while, intelligible problems.

5. Textbook makers are apparently coming to appreciate the need for making their problems representative of the fields of activity in which pupils are likely to be engaged.

6. Movement in all these directions toward better third-grade teaching material is as yet, however, ill defined and irregular, resting on no firm basis of educational theory. The pressing need of the present time is for a pedagogically sound definition of arithmetic material (a) in terms of the amount needed to accomplish most economically the desired results, and (b) in terms of subject-matter looking not alone to efficient mastery of the fundamentals but to the proper development of the whole child. Until we have such a definition, the making of textbooks in arithmetic must be (as it has been in the past) guesswork, pure and simple.

Editorials

FINANCIAL RESEARCH

There has been great progress in the administration of our city school systems during the past ten years due to the introduction of more adequate child accounting. In every progressive school system data with respect to retardation and elimination, together with the measurements of the achievements of children are being used as the basis for revision of courses of study and the development of a more adequate organization of schools.

One may question, however, whether any comparable advance has been made in the fiscal administration of schools during the same period. As one reads even the more recent legislation with respect to the distribution of school funds one cannot but be impressed with the unscientific methods which are employed in most of our states. If we accept the ideal of equality of opportunity, we certainly have not yet been willing to fight for its realization in terms of an equitable development of grants-in-aid.

In our city school system there is still need for much more adequate accounting than is commonly found. In most American cities the preparation of the budget and the actual fiscal administration of the school system are less adequate than that commonly found in successful business organizations. If our system of financial accounting permitted the comparison of costs among the several school units making up our city school systems, and if we were able, as well, to report to our public the cost of teaching English, arithmetic, the social sciences, or the cost of kindergartens, of health service, the household arts, and the like, our position in these days of retrenchment would be much more secure.

There is likewise need for a most careful scrutiny of the methods employed in financing building programs throughout the United States. Large expenditures are being undertaken in some cases without an adequate study of increases and shifts in population within the area to be served; and in some cases extravagances have been permitted in the construction of school buildings, which could not possibly be justified were the program of capital expendi-

tures for a period of years put clearly before the boards of education and the public.

We need in state, county, and city school systems directors of research whose main purpose will be the study of the fiscal problem. One can hardly expect that those who are primarily concerned with child accounting, organization, the achievements of pupils, and methods of learning can find the time, nor that they will be equally expert in dealing with the problems of fiscal administration in our school system. It will be equally futile to hope for thorough-going scientific work in this field from bookkeepers and purchasing agents without scientific training. No phase of research that has been undertaken promises more certain returns in increased efficiency and in economy in the administration of public education.

G. D. S.

A GENTLE SUGGESTION

In any plan for training teachers in service, acquaintance with the best in professional writing should be included. A recognition of this is afforded by the development of the reading circle idea. Certain books, adopted for inclusion in the reading circle list, have undoubtedly been widely read. Presumably, they have also influenced practice.

It is well known, however, that a great many of the best suggestions for teachers are found in periodicals rather than in books. To go no further than the pages of our own journal, Dalman's concrete example in the January, 1920 number of differentiated requirements for pupils of varying abilities remains the best answer with which we are familiar to a question which thousands of teachers are asking, namely, "After grouping children on the basis of ability, how shall I go about making a course of study for each group?" Nevertheless, it is clear that Dalman's article, though a great little idea, can't furnish forth a book. Accordingly, it can't be included in the reading circle material.

Again, we know of no more satisfactory analysis of fractions than Kallom offered in the March, 1920 number. He gave a test for proficiency in each type of addition of fractions and the form on which the record was made. This form, when made out, shows the types of examples in which the pupil has difficulty. Teaching and drill on these types follow; and the pupil is then

given another form of the test. He is excused from further work in addition of fractions when he attains a consistently perfect score on one of these tests. Here we have curriculum analysis, diagnosis, individual instruction, and tested results—all organized in a highly satisfactory manner. Yet the teacher must remain unfamiliar with this admirable presentation because it does not appear in a book.

At the risk of being tedious we shall give another instance. Doctor Luella M. Pressey s suggestive article in the September, 1920 number must remain unknown to the vast majority of the teachers who could use it. Yet in this article Doctor Pressey not only adds an important test to those already available (and just when we needed it, too), but she also sets forth a method by which the teacher may construct home-made objective tests of her own. At some future time the book writer—whose service consists for the most part in collecting other people's ideas, and whose only contribution may be his particular juxtaposition of them—will no doubt have something to say about this test. He may even refer to Doctor Pressey's article. But it is certain that his write-up will lack the directness, the force, and the vividness of the original. Moreover, it will be late in arriving upon the scene. If an idea is worth using, it is worth using now.

The worth of these ideas does not depend upon their appearance in a bound book. What Rochester is doing in reading (O'Hern and Hawley), what Boston is doing in geography (Barthelmess), what St. Louis is doing in penmanship (Walker), what the Pacific coast school is doing with intelligence tests (Terman, Proctor, Dickson)—these are among the things which wide-awake teachers want to know and which wide-awake superintendents want their teachers to know.

Why, then, may not subscriptions to periodicals or single numbers of periodicals be included in the material which teachers are encouraged or expected to read? Why this predilection for books? In our judgment there is in any one of a half dozen magazines more timely and stimulating educational writing than can be found in any but the most unusual book.

B. R. B.

Reviews and Abstracts

LEWIS, E. E. *Scales for measuring special types of English composition.*

Mr. Lewis has done a painstaking piece of work and has produced in this monograph several new extensions of the original Hillegas Scale. Judges were carefully selected and trained. Consequently one feels that the values attached to the specimens have a greater degree of reliability than is usually the case. Teachers of English composition will no doubt welcome these scales not only for these reasons but also because of the feeling of ease with which letters can be evaluated in terms of the scale.

While these scales do not meet all the standards set up by the author himself, nevertheless their shortcomings—such as unequal steps between specimens, a single specimen at each step, and lack of distinction between form and content values—are not serious enough to raise a valid objection to their use.

Although no evidence is given to show that letters can be more accurately scored with these scales than with either the Thorndike Extension or the Nassau County Supplement of the Hillegas Scale, it is not improbable that this will be found to be true. On the other hand, had the group of judgments been split into two parts and the scale values found on the basis of the two groups of judgments separately, it would undoubtedly have been evident that the very high coefficients of correlation give an exaggerated impression of the stability of the value assigned to the specimens. In the scales themselves, however, the dropping of the second decimal place tends to correct this impression.

In constructing the scales both the method originally employed by Hillegas and in part the method used by Trabue in deriving the Nassau County Supplement, were tried. A somewhat detailed examination of the specimens of social letters and of narration shows that the two sets of values for the latter differ on the average by 0.26 of a step, or about 1/32 of the length of the scale, while for the former the average difference is only 0.17 of a step. For the specimens of social letters the positive and negative differences appear in small groups; in the narration specimens they appear in very large groups. This seems to be due to the fact that the very small differences between the specimens were added together to get the total scale length, whereas in the original scale larger and hence more reliable differences were added together. A very few errors entering in this way seem to account for the larger difference between the two sets of values in the narration specimens.

In view of this condition the values obtained by the shorter method of having the specimens graded is probably more accurate than those obtained by the much more involved process. The work of the author in using the original method has, however, been by no means unprofitable; for, as a result of it, the stability of the original Hillegas values has been further substantiated. Moreover, the method of having specimens graded by competent judges with a scale has definitely established the possibility of obtaining various extensions of the original scale and of securing steps at equal intervals.

M. J. VAN WAGENEN

University of Minnesota

HERTZOG, W. S. *State maintenance for teachers in training.* Baltimore: Warwick & York, Inc., 1921. 144 pp.

The author, at the outset, portrays with desirable emphasis the fundamental importance of relieving the teaching profession of an unfavorable social attitude, and the necessity of affording substantial economic relief to competent members of the profession. This is not new; but until these cardinal defects are remedied the teaching profession will continue to occupy a position of inferiority with respect to other professions and with respect to the industries.

Among the purposes laid down in this volume are the following: (a) To investigate teacher-shortage and to record the best relief measures employed in several states, (b) to investigate the principles, problems, and practices involved in a system of subsidies for prospective teachers as one method of recruiting the profession. These purposes are detailed, (a) by a survey of conditions which reveal an imperative need for financial support to teachers in training, (b) by a review of efforts in foreign countries and in the United States to recruit the profession by means of subsidies to prospective teachers, and (c) by drawing lessons from other professions and the industries in their attempt to win recruits by means of financial assistance.

The author concludes that a policy of subsidies, following the principles and practices which obtain in the recognized professions, would contribute substantially to the relief of the existing teacher-shortage and would in large measure raise the profession out of the slough of social inferiority. The volume merits a careful reading by all those who are vitally concerned with teacher-training and by school administrators who have been far too indulgent in accepting the "finished products" of many of our teacher training agencies.

Your reviewer would have welcomed a discussion of the significance of a teacher-shortage as an opportunity on the part of school administrators to impress the public with the dangers involved in an inadequate supply of well-trained teachers. School administrators appear reluctant to place the responsibility for a shortage of competent teachers where it belongs, namely, on the public. They have lowered standards and made possible the employment of thousands of untrained and incompetent teachers. Report has it that 48 percent of the teachers in one state of our glorious Union have had no training beyond the eighth grade, and that 68 percent have received less than a high-school education. If state laws permit this disgraceful condition, the obvious remedy is the establishment of standards. In states maintaining desirable standards, the practice has been to lower them. Such practice reduces the shortage by opening the gates to incompetents, and accustoms an undiscriminating public to a low standard of teaching.

Nothing could be more wholesome to the profession and more beneficial to the nation than a rigid enforcement of standards entailing consequences for which the public should be held responsible. So long as school administrators are unwilling to place the responsibility on the public, where it rightfully belongs, but are willing to employ incompetents in periods of teacher-shortage, so long will the profession endure economic servitude and occupy a position of social inferiority.

GEORGE F. ARPS

Ohio State University

AVERILL, LAWRENCE AUGUSTUS. *Psychology for normal schools*, (Riverside Textbooks in Education). New York: Houghton Mifflin Company, 1921. 362 pp.

Unfortunately for the training of teachers, mere contact with courses in education, no matter what they are, is in many quarters still supposed miraculously to give professional preparation for teaching. Legal requirements for certification often merely demand a certain number of hours in psychology and education, regardless of whether the courses serve a specific practical purpose or meet specific needs. Similarly, textbooks on education are like the old time physician's shot-gun prescriptions, being labelled as suitable for college or normal school classes or for teachers' reading circles. A textbook that is suitable for groups so diverse in preparation, interests and needs is probably not good for much.

Professor Averill's book has the merit of being designed for a specific group, bearing as it does the title *Psychology for Normal Schools*. It is not a text in educational psychology, but in general psychology, though it includes much of what is ordinarily treated in educational or genetic psychology. On the other hand, it differs very much from the ordinary text in general psychology. The first one hundred and forty-three pages, or twenty-one lessons out of forty-six, are given to a detailed discussion of instinctive and emotional behavior, followed by forty-five pages (six lessons) on heredity. Sensation gets seven pages with little in them but a doubtful complete list of sensations and four pages on the earliest sensations of infancy. Six pages dispose of perception, while the whole subject of the perception of space and time gets twenty lines. Memory is discussed in one lesson of six pages without a word about the laws of learning which are so admirably presented in such a book, for example, as Pillsbury's *Essentials of Psychology*. One lesson likewise disposes of such an important topic as thinking and another of the will and moral development. The remaining lessons deal with the juvenile delinquent, the subnormal child, the gifted child, individual difference, the unstable child, adolescence, and the evolution of the social attitude toward children.

What psychology is of most worth for teachers is a question that will be variously answered. That the type of psychology presented in this book will be accepted for normal schools as better than the conventional type the reviewer doubts very much. The student would complete the course outlined with very little notion of what modern psychology is and without and appreciation of the complexities of mental processes. Nor does it appear that he would be better prepared for teaching by it. Surely to give one lesson to a discussion of the food-getting and hunting responses and only one each to such topics as sensation, memory, and thinking is out of all proportion.

V. A. C. HENMON

University of Wisconsin

STEVENSON, JOHN ALFORD. *Project of teaching*. New York: Macmillan Company, 1921. xvi+305 pp.

This book is a valuable addition to the growing literature which attempts a systematic examination of the proper definition of the term "Project"; and a constructive, though critical, evaluation of the project method as a practical instrument for school use. Too much of what has been written on this topic either takes the form of propaganda or is so theoretical that it has little stimulation for the average school man or woman. Both of these weaknesses are avoided by Doctor Stevenson. The issues involved are, in general, put clearly and fairly, and are made concrete by considerable

illustrative discussion. The book should be welcome, therefore, as a text even by those who would object to certain of the author's conclusions. Those who have followed most closely the development of educational practice in the last twenty years, will probably raise one or more of the following objections to the treatment of the project as found in this book.

(1) The four standards set up by Stevenson must not be regarded, by implication, as standards for judging teaching in general, but only as standards for judging project teaching. Some will feel, moreover, that these standards do not represent an accurate digest of present and past opinion regarding the project, but rather the author's choice of certain elements among these opinions. Such a selection is, of course, justifiable, but the reader must keep in mind that such a choice has been made.

(2) The book should have contained a section treating the practice of schools which the teaching public has thought of as employing the project method. Certainly there should have been a critical review of the work of Booker T. Washington, at Tuskegee; of Dewey, at Chicago; of Miss Cooke, at Francis Parker; of Bonser, at Speyer School; and of Meriam, at Missouri. Possibly, too, some of the more radical schools—such as that of Miss Johnson at Fairhope—should have been evaluated. In other words, many readers will doubtless feel that there has been too much attention given to those who have written about the project, to the neglect of those who have contributed in practice to the development of the various technics involved.

(3) The various issues involved should have been made clear by a careful review of the chief theories which have been prominent in the last quarter of a century, particularly the theories which have gone under the following names: Herbartianism, Froebelianism, pragmatism, reaction psychology, and the problem method as expounded by Dewey. No doubt the author was led to omit the discussion of all these theories except the last, on account of the difficulty of presenting a clear discussion within the limits of a book of the character of this one.

(4) The distinction between the psychology of thinking and the psychology of learning is not made sufficiently clear. Most of the mistakes which have been made in connection with the problem method and the project method may be traced to this confusion. We learn, to be sure, while thinking, but this is not the only, nor always the most economical method of learning.

Some of these objections are referred to, or touched upon, in various places throughout the book. No doubt with more time and space, all of them would have been more adequately discussed.

The author is to be especially congratulated on his steady refusal to be led astray by the recent emphasis on the more subjective aspects of the project; and on his consistent avoidance in his illustrations and discussions of the sentimentalities in which too much of our recent literature abounds. He is not seeking a method which mysteriously develops certain subtle and general qualities. He seeks a method through which plain objectives in the form of special abilities of known value may be most economically reached. He is undoubtedly right in his statement that "the provision for the natural setting of the teaching situation is the distinct contribution of the project method." In fact, the elaboration of this idea is the chief contribution of this book.

Ernest Horn

State University of Iowa

SNEDDEN, DAVID. *Sociological determination of objectives in education.* Philadelphia: J. B. Lippincott Company, 1921. 322 pp.

In this book Doctor Snedden takes the position that the objectives of education must be determined, in the main, by data and methods that fall within the field of sociology. In order to equip the youth for the business of living, it is necessary to furnish training for the various occupations of adult life. Such training must be directed by studies of a sociological sort. Moreover, sociology can inform us as to the value or function of educational subject matter which has no direct bearing on vocation. "By the application of suitable sociological methods," for example, "it is entirely practicable to discover the scope and character of mathematical knowledge now used in any given 'standard of living' class or group, and, on the basis of the facts thus found and evaluated, to propose necessary or desirable improvements in processes of instruction and training to be applied to the rising generation" (p. 33).

In Doctor Snedden's exposition, this doctrine is linked up with the distinction between 'values of production' and 'values of consumption.' The former values have to do with vocation, and hence the standard is set by the needs and demands of society. The latter, on the other hand, "center definitely in cultivation of specific personal, intellectual, and aesthetic interests—the resources wherewith we enrich our leisure time, our individual lives. . . . They should at least establish abiding cultural interests—appreciations, tastes, enthusiasms, even hobbies—in literature, science, foreign languages, and history" (p. 83). This distinction in values points to a corresponding distinction in educational subjects or courses. One group, labelled "A class" subjects, "should lead very directly to processes and capacities known to be of use to the individual in adult life, or to the society of which he shall be a part" (p. 70). The second group, called "B class" subjects, "are those primarily which we follow, and to the degrees only which we follow them, because of innate and easily stimulated desire" (p. 71). The subjects in the former group are "hard"; they are properly imposed only because of obvious usefulness, and they are characterized by rigid method and by drills and tasks which are aimed at doing, at facility of expression in action. The subjects of the latter group, on the contrary, are directed rather towards absorption, or assimilation (p. 49), and so they partake more of the nature of "high-grade play."

It is this distinction between values of production and values of consumption that constitutes the basic principle of Doctor Snedden's philosophy of education. From this distinction he deduces the classification of subject matter into "A class" subjects and "B class" subjects, and also the further contrast between hard work or drill and "high-grade play." The former aims at "expression in action," while the latter is directed towards "assimilation" or "absorption." It is rather curious to note that in this exposition practically no account is taken of that considerable body of recent psychology which converges upon the conclusion that all knowing is doing, and which, accordingly, would repudiate the distinction in question. The trend of Doctor Snedden's doctrine at this point is not forward, but backward, toward the Lockean conception of the mind as just a wax tablet or sheet of white paper, upon which the environment may write what it will. Perhaps his position is defensible, but the psychology which underlies it has been too much under fire of late years to be taken for granted without argument.

But this is not the only point at which Dr. Snedden leans heavily upon a dubious tradition. His distinction between vocational and non-vocational or cultural subjects

is another instance of the same thing. The distinction is plausible only as long as we assume the validity of the historic opposition between vocation and culture. In spirit or emphasis Doctor Snedden's doctrine does not differ materially from the familiar notion that business or vocation is a thing quite apart from the nurture of the soul. The distinction presumably has a certain validity if we think of culture as, in the main, a private or subjective affair, in that it is directed towards the cultivation of detached, inner appreciations, which have no significant connections with the world of affairs. If we accept Dewey's definition of culture as "the capacity for constantly expanding in range and accuracy, one's perception of meanings," the opposition disappears and with it the hard and fast classification of subject matter into "A class" subjects and "B class" subjects. The efficacy of a plumber, for example, is determined in part, no doubt, by his training in the specific activities that constitute plumbing, but only in part. It is likewise determined by his conception of the relation that should obtain between employer and employee, by his standards of honesty, by his aesthetic appreciation, in short by his ability to see the meaning of his vocation thru its relations with a wide social context. To classify subjects offhand as vocational or non-vocational is to encourage the spirit of narrow vocationalism and to give aid and comfort to the exponents of an equally narrow and abstract "culture," against which Doctor Snedden so emphatically registers his opposition.

One further point should be noted. For the determination of educational objectives, the subject of sociology is, to Doctor Snedden, as the cloud by day and the pillar of fire by night. As far as vocational subjects are concerned, this view is plausible enough, provided that we take the meaning of vocation in a sufficiently narrow sense. Taken in this way, we can ascertain what is required for a given vocation as objectively and impersonally as we can determine the constitution of water or calculate the distance from the earth to the moon. But if we conceive of vocational training as something more than fitting a man to a groove, so that he can go through life with a minimum of intelligence, the matter takes on a different aspect. Doctor Snedden himself implies this when he says, in connection with non-vocational subjects, that the curriculum must be constructed "on the basis of the facts thus found *and evaluated*," but the significance of the evaluating evidently escapes his attention. He seems to take for granted that the evaluation of a more or less incidental task, which sociology is entirely willing to take off our hands and thus relieve us of all responsibility in the matter.

Given the same sociological method, there would presumably be no difference in the objectives of such men as Bismarck and Lincoln. Doctor Snedden gives no hint that the determination of objectives differs in any important respect from the process of prospecting for oil or coal. To be sure, the objectives are not supposed to be standing around in obscure places, waiting to be discovered, but there are facts to be discovered, and these facts, when brought to light, decide the question of objectives. Perhaps this too may be conceded. But can we assume that different individuals will discover the same method? Are we, or are we not, obliged to make allowance for a personal factor which somehow falls outside the field in which the sciences ply their trade, and which, because it tips the balance in favor of accepting certain facts as facts, has the last word in the determination of objectives?

This does not mean, of course, that the selection of objectives is to be left to random choice or caprice. But it does mean that the adoption of a gospel of brotherly love, of caring for the weak and of subordinating personal interests to group interests, or the adoption of any other gospel, involves certain factors which inevitably introduce

individual differences. Different men are bound to attach different meanings to the experience of freedom and responsibility, and to the desirability of cultivating an attitude of broad tolerance and sympathetic understanding, which means that they are bound to report differently upon the facts of experience. No scientific method, however rigorous, can provide an escape from these individual differences. The only rational check that can be applied lies in the organization of the facts which we happen to recognize into a coherent system, which then constitutes our philosophy of life and which, on occasion, becomes the basis for our determination of the objectives of education. It is simply self-deception to assume that in the determination of objectives the investigator can keep his deepest, nonscientific self out of the situation. To make such an assumption is scientific method gone mad.

It is easy to understand the present-day reaction against philosophy in its relation to education. In the past philosophy has served only too often as a substitute for the painstaking accumulation of fact by which educational theory and practice must be informed and directed. Perhaps Doctor Snedden is right when he says that "the philosopher can not, of course, be very patient of these attempts thus to consider democracy analytically." (p. 290). If I may trust my own observation, however, he is less disposed to be irritated by analysis than by the lack of it. The classification of values proposed by Doctor Snedden, the contrast between expression and absorption, and the distinction between work and play, are all sadly lacking in serious analysis. They are all conceptions that have had a long and evil history, and their weaknesses have been pointed out on many occasions and by many critics. Instead of substituting a more profound analysis, Doctor Snedden revamps these old conceptions and offers us a questionable philosophy of education, thinly disguised as sociology. I am glad of the opportunity to express my sincere admiration for the clearness and suggestiveness of Dr. Snedden's exposition; but I find myself wholly unable to agree with him as to the validity of his method and of his conclusions.

B. H. Bode

Ohio State University

News Items and Communications

This department will contain news items regarding research workers and their activities. It will also serve as a clearing house for more formal communications on similar topics, preferably of not more than five hundred words. These communications will be printed over the signatures of the authors.

Something from Minneapolis

According to Assistant Superintendent W. F. Webster, Minneapolis is about to establish a Bureau of Research. It is evident, however, that without a bureau organization Minneapolis has already been doing some of the work which a bureau ordinarily undertakes. Readers will be interested, for example, in Superintendent Webster's article in the current number of "The League Scrip." It is entitled "A Statistical Story." The author has at his disposal the records of pupils admitted to each of the various grades

for the past twenty years. His opportunity to study elimination is therefore unusual. It is not often true that reliable educational data are available for historical treatment. When, as in this case, the data include not only enrollments but numbers retarded, numbers promoted, and facts as to cost, the opportunity for significant reporting is obvious.

Research in Teacher Training

President H. A. Brown of the Oshkosh Normal School is evidently concerned because the light of investigation is not being turned on problems of teacher training as often or as fully as it should be. We agree with him. Anything that can be done in the selection and training of teachers is done at the very focal point of our educational enterprise. It is certain that the training of teachers is a topic on which no one can be dogmatic. The very diversity of the types and amounts of training indicates the lack of agreement. There must be best ways in which teachers may be trained. The field of teacher training is a promising one and we feel sure that any research worker who enters it will feel amply repaid for having done so.

Research in Higher Education

We are receiving a number of letters and reports which indicate that educational research is being applied to higher educational institutions. There is every reason why this should be so. The problems are scarcely less numerous that they are with reference to elementary and secondary education. The treatment of these problems, however, has hitherto been largely characterized by prejudice and an absence of objectivity. It seems to be especially true that the Bureaus of Educational Research which have been established in universities are being called upon to turn their attention to the condition of the institutions to which they belong. Perhaps research like charity may well begin at home.

Graduate Work and Salary Schedules

A recent correspondent raises the question of what is being done in the public schools to encourage teachers to pursue advanced work subsequent to graduation from a college or university. In other words, the question is whether it is possible for a teacher who has received training additional to that required for the baccalaureate degree to receive a higher salary or more rapid promotion, than the teacher who has not had this additional training. We shall be glad to receive information from cities whose salary schedules or whose regulations concerning promotion take into account the question of graduate work. It is our belief that account should be taken of it.

The Department at Okmulgee

Miss Velda Bamesberger, Director of the Department of Educational Statistics at Okmulgee, Oklahoma, has an organization which seems well adapted to serve the schools of a moderate sized city. Her department consists of three full-time people—a psychologist, a general assistant, and herself. It is surprising how much a well-organized department of this size can do if it succeeds in enlisting the support of the teachers—as Miss Bamesberger has apparently done. She says in a recent letter: "I have been devoting a great deal of my time to devising records, etc., for the system, and have also completely classified the seventh and eighth grades on the basis of intelligence tests, educational tests, and school marks. Our recent educational tests show that the classification is working well, and the high-school principal and teachers are more than pleased with it."

A Professional Duty

In the November number of the Detroit Educational Bulletin and under the above caption, Superintendent Frank Cody of the Detroit public schools presents vigorous plea to his teachers to join the National Education Association. After reviewing clearly and concisely the change that has taken place in the last four years in the public attitude toward education in general and toward teachers in particular, he shows that this change was largely brought about through the work of the National Association. He further points out that only if every teacher is solidly behind the N.E.A. can that Association furnish the leadership necessary to secure adequate public support for the schools of today and for the still better schools of tomorrow.

The article is illustrated (if this term is properly used) by the following quotation from Andrew Carnegie, which is printed in bold face, old English type: "You cannot push a person up a ladder unless he is willing to climb a little himself."

Bucyrus Survey Bulletin

Superintendent J. R. Patterson of Bucyrus, Ohio, has sent us one of the most pretentious survey bulletins which we have received. It is number one of volume three of his series. Superintendent Patterson has been conducting a cycle of three surveys—fall, mid-year, and spring, during the past two years; and the present bulletin gives data on the contents of the program. The bulletin is designed for the information of his teaching and supervisory staff and to disseminate survey notions among local citizens and interested fellow workers elsewhere. The contents reveal explanatory notes on statistics sufficient to enable the teacher, untrained in this line, to understand the meaning of the terms used in the body of the discussion.

The report of children's abilities involves column spelling from the Ayres-Buckingham scale, Courtis, Woody-McCall, and Monroe arithmetic, English Composition, Monroe reading, and hand-writing scored by the Ayres scale. There is also a section on age-grade conditions. It is unfortunate that, like so many others, Superintendent Patterson confuses retardation with overageness and acceleration with underageness, though his final statement to his last chart reveals the fact that he recognizes the real meaning of an age-grade table.

We hope that Superintendent Patterson will favor us with a copy of his survey bulletin, Vol. III, No. 2 when he completes it. We shall look forward with interest to the results contained therein and the treatment of the data which he will make available.

Another Survey Bulletin

There has also come to us a Survey Bulletin from the Kent Ohio city schools reporting the use of standardized tests during the second semester of the past year. The testing program here was somewhat more ambitious than the one at Bucyrus, since general intelligence tests as well as achievement tests were used, and the measurement of achievement involved a larger number of subjects. The subjects in which measurement was made were spelling, writing, arithmetic, composition, silent reading, geography, history, music, first-year algebra, second-year geometry, and third-year physics.

The bulletin, like the one previously described, includes numerous tables and graphs together with pointed comments on the scores and on the remedial measure which will be undertaken.

We have just received from W. H. Pillsbury of the Buffalo (N. Y.) city schools, a mimeographed copy of the "Buffalo Teachers' Library." This not only lists what

The Buffalo Teachers' Library was thought by the committee of Buffalo teachers and administrators to be the best books and magazines for the teachers but also gives a brief review of each book. A sentence or two which shows the main problem and method of attack makes the list more valuable than the name of the book alone. The material has been classified under the following heads: (a) general reading; (b) elementary subjects; (c) secondary subjects; (d) vocational schools; (e) intermediate schools; (f) extension work; (g) teachers' organizations; (h) magazines for teachers; and (i) book companies. The first four of these large divisions is further subdivided and the material appropriately classified.

The work of compiling this library was done some two years ago, but lack of funds prevented it being printed at that time. It was mimeographed in order that it might be placed in the hands of the teachers. In his letter to us Mr. Pillsbury made the following statement: "We propose to revise it from time to time and are putting it out now with the hope that we will get suggestions and changes which will make each edition a little improvement on the preceding one. Any suggestions which may occur to you in the way of eliminations, additions, substitutions, or changes of method of preparation of the list, will certainly be most gratefully received."

The list, as it stands, is valuable to any of our readers. We suggest that superintendents of schools and research directors write Mr. Pillsbury for a copy, and having received it, show their appreciation by sending to him such suggestions as may occur to them.

English in Secondary Schools A Report on the Conditions of the Teaching of English in the Secondary Schools of New Jersey (published by the New Jersey Association of Teachers of English and obtainable from the Secretary, Miss Mabel A. Tuttle, Linden, N. J.) will be of interest to all those concerned with high-school administration for two reasons: First, it presents in small compass a rather complete survey of the English situation in a state, treating details of schedule, curriculum, methods, etc. Second, it presents most frankly the attitude of the English teachers themselves on these questions.

In addition, the form of the report lends itself to ready comprehension. A short summary gives a bird's eye view of the situation, then each point is explained by numerous representative quotations from the questionnaires on which the report is based, and finally a schedule of eighteen "Recommendations" gives very definitely the standards and conditions which the teachers as a body think should prevail in high-school English teaching. With so much dissatisfaction abroad in regard to the results of English work, all those concerned should take this opportunity of securing further insight into existing difficulties and of learning what remedies an active body of teachers have tried.

E. W. Dolch, Jr.

University of Illinois

A Different Kind of Report Card

Superintendent E. B. Sellew of Middletown, Connecticut, has furnished us with a copy of his special report to the Board of Education and to parents of pupils in the Middletown schools. Since it is a radical departure from the usual form of report card or leaflet, it was thought that our readers will be interested in it.

The report is a four-page folder. The first page contains a black space for the name of the parent to whom the report is being sent and a general explanation of the

importance of school work with a plea for cooperation. Page two lists the subjects taught in high school together with the total number in each class and the number making the various grades or marks. For example, English I B, which means the first semester of first-year English, had 248 members, 47 of whom received grades of 90 to 100, 49 received 85 to 89, 105 received 75 to 84, 37 received 70 to 74, 10 received 60 to 69, and none received below 60%.

Grades of 85 to 100 are designed as honorary grades while any grade below 70 is a failing grade. The group in which the child is located is indicated by a red line. This calls the attention of the parent vividly to the number of students in the same grade with his child and to the number who received marks higher or lower than his child received.

Pages three and four give additional information concerning the schools, page four showing the school census by grades and the residence of high-school pupils with percent who are paying tuition.

While the report may be somewhat difficult for many of the parents to read, it is worth considering because of the large amount of information which is thus placed in the hands of patrons of the school.

A Few Data on the Use of the Stanford Revision of the Binet-Simon Tests by Halves

Apropos of the article in the September issue of this journal by Messrs. Otis and Knollin, the following data are submitted upon some of the same points covered by the article. In the course of several years of testing the writer accumulated a number of complete records of tests of pupils and became interested in making a comparison between the pupils' mental ages computed from the first and second halves of the tests. That is, each pupil's test record was divided into two parts, one representing his performance upon the first three or four, as the case may be, of the tests for each mental age level and the other his mental age as based upon the last three or four tests for each age. In doing this each test was, of course, given weight for just twice the number of months given when all the tests are used together. At the time of testing the pupils whose records are here treated, the writer had no thought of making such a study of their records and hence the giving of the tests could have been in no way affected by such a plan.

The records of 182 pupils were complete so that they could be included in this study. These pupils were well distributed through all the eight elementary grades and had mental ages ranging from four years and eight months to seventeen years and two months. After dividing their records into first and last halves by years as described above, the ages by the first half range from four years to sixteen years two months and by the second half from five years four months to eighteen years two months. The median age by the second half is 2.8 months higher than that by the first half. The average is 4.8 months higher. This latter figure seems to be unduly affected by a few rather freakish cases most of which occur at the mental ages of four and five and eleven and twelve. Except for these there is no apparent tendency for the age level of the pupils to make very much difference in their comparative mental ages by the first and second halves of the tests.

The following table gives the total distribution of differences in ages between the two halves. Positive differences mean that the ages by the second half are greater.

DIFFERENCES BETWEEN MENTAL AGES BY THE FIRST AND SECOND HALVES OF THE TESTS

Months	Number of Cases
−24–21	1
−20–17	1
−16–13	0
−12– 9	5
− 8– 5	17
− 4– 1	25
0	23
+ 1– 4	33
+ 5– 8	26
+ 9–12	22
+13–16	10
+17–20	8
+21–24	4
+25–28	1
+29–32	2
+33–36	2
+37–40	0
+41–44	1
+45–48	1
Total	182

As is stated above the median difference when signs are taken into consideration is 2.8 months. Neglecting signs and merely considering the absolute value of the differences, the median difference is 5.5 months. The average absolute difference is 8.2 months, being made much larger than the median by a few rather extreme cases. It will be noted that the seven greatest differences are all differences by which the second half gives a greater age than the first half. Also these seven are scattered amongst six different age levels, so it is apparent that the tests at any one age are not at fault here. The unusual results from the tests at ages four and five and eleven and twelve which were referred to above were not due to a few such extreme differences but to a larger number of differences of from eight to sixteen months.

The product-moment coefficient of correlation between the ages by the two halves is .92 with a P. E. of .01. The probable error of measurement[1] is found to be 5.5 months. This agrees very closely with the median absolute difference of 5.5 months mentioned above, and also is in substantial agreement with the results found by Otis and Knollin. This probable error of measurement is between four and five percent of the average mental age. This may be compared with seven percent which is the smallest probable error of measurement of any group intelligence test that the writer has noted.

C. W. Odell

University of Illinois

[1] The Probable Error of Measurement is obtained by the formula $P.E.M. = .6745\sigma\sqrt{1-r}$. For σ the average of the two standard deviations is used. r is the coefficient of reliability, that is, the coefficient of correlation between the first half and second half.

National Association of Directors of Educational Research

(E. J. ASHBAUGH, *Secretary and Editor*)

Grand Rapids, Michigan.—Charles D. Dawson, Assistant Superintendent in charge of Research, has sent us a study of the marks given high-school pupils for the second semester of the present year. Grand Rapids has just changed its marking system from a percent basis to a five-letter basis. A rather extensive argument is included showing that the whole system of marking passing pupils between 75% and 100% is unscientific because the best and the poorest pupils often differ by more than 25% and because the different marks within these 25 units were not evenly distributed.

The new system gives the marks, A, B, C, D, and E and "Condition" with percentages of 10.5, 30.0, 30.0, 17.0, 9.5, and 3.0 respectively. This shows the marks to be skewed toward the upper end with the exception of the E value. Since the extent of this skewness increases from the 9th to 12th grades, it is held that the marks agree in practice with the theory that the successive grades in the high school represent greater and greater selection.

A great deal of variation is shown in the distribution of marks in the various subjects. It is stated, that "This is due mainly to variations in the type of students of which the different classes are made up, to the personal equation of the teachers who do the marking, and to the degree of difficulties of the work necessary to earn a passing mark in the different subjects." It is suggested that the teachers examine carefully the group of failures in their subjects as well as confer with one another concerning the use of the scale.

Those interested in making a study of their own distribution of marks, might well write Mr. Dawson for a copy of this report.

West Allis, Wisconsin.—T. L. Torgerson, Director of the Department of Educational Measurements, has devoted Bulletin No. II of his department to tests and measurements. He states that the bulletin is issued to give the new teachers a general idea of the scope and aims in the field of tests and measurements. A statement is made relative to general intelligence and some fundamental terms, definitions and underlying principles are set forth. The following statement is underscored for emphasis and is doubtless intended to be the key note in the measurement work for this year: "The pupils of the West Allis schools this year will be measured in terms of their own ability. The goal or standard for each pupil to reach is an achievement approximating 100% in efficiency."

His review of the work last year shows measurement in spelling, silent-reading, composition, arithmetic, and algebra. Statements are made regarding methods used, results obtained, and reaction of teachers to remedial work. Extensive intelligence testing was also done, and the following statement is made concerning the utilizing of results:

1. Mental age of the pupil used as a basis for grade placement.
2. Intelligence quotient of the pupil used as a basis for classification.

3. Intelligence and achievement compared to determine the achievement quotient of each pupil.

4. Aid for individual diagnosis.

5. Permanent record for future diagnostic purposes.

6. Discovery of sub-normal pupils.

A tentative outline of work for the year 1921-1922 includes for the high school a testing program in geography, history, and algebra; a testing program for the improvement of spelling, silent-reading, arithmetic, and handwriting; diagnosis of Clapp's Language test; a prognostic test of mathematical ability; a study of the vocabulary of the high-school pupils; testing for the appreciation of poetry in the junior high school; one group intelligence test and a study of the distribution of marks.

In the elementary schools a program of testing for remedial work in spelling, silent-reading, arithmetic, handwriting, language, and oral reading will be conducted. There will also be intelligence testing, a study of school marks, and a continuation of the study of the promotion of pupils on the basis of mental age. Closer supervision of the special help period for individual instruction will be given.

Denver.—Miss Emma M. Brown, Director of the Department of Measurements and Standards, has recently sent us two bulletins. Bulletin No. 2 deals with daily programs. Summarized, the general conclusions reached after a study of the teachers' daily programs are as follows:

1. There is a lack of uniformity as to accuracy, distribution of subjects, free time, and number of pupils.

2. Due to the fact that programs are not made accurately and carefully, there is much confusion as to interpretation.

3. A study of the principles involved would be helpful alike to principals and teachers.

4. Supervision of program making by the principal tends effectively to unify the work of the school in the following ways: (a) distribution of teachers' load as to teaching time, number of pupils, and free time; (b) distribution of pupils' time among various subjects; (c) reduction in amount of unsupervised study; (d) provision for coaching pupils who have been absent, or for helping backward or accelerated pupils; (e) variation in course of study to meet local needs; (f) provision for mental and educational testing and guidance; and (g) a more scientific placement of subjects. The tabular material presents data from which these conclusions were drawn. The whole bulletin is suggestive to those who wish to improve the daily programs of teachers in their school systems.

Bulletin No. 3 reports two tests in spelling in elementary and junior high schools, grades II-A to VIII-A inclusive, September 1920 and January 1921. The bulletin not only presents in figures and graphs the conditions in the various Denver Schools, and in the city as a whole, but also includes deductions, suggestions for improvement, and some general principles for the benefit of the teachers.

The following are some of the results:

(1) That girls spelled better than boys; (2) that there was a difference between the ability in content and list spelling which gradually decreased in the higher grades; (3) that locality of school and size of class had little effect; (4) that schools with foreign populations were handicapped in the lower grades, but overcame the handicap in the upper grades.

Under suggestions for improvement, the following statements, while not new, will bear repetition:

1. Secure automatic accuracy for words of greatest frequency in school work, social usage, and adult correspondence.

2. Give preliminary tests in order that time may not be wasted in teaching children words which they already know.

3. It is more important that a pupil should have a definite and usable method of learning than that a high technique should be developed on the part of the teacher.

4. "Trouble spots" occur in some words and not in others and frequently change in the same word from one grade to another. For this reason it is doubtful if it is worth while to spend much time in calling attention to them.

5. Homonyms may be effectively taught separately in initial presentation and together when words have been confused.

Kansas City, Missouri.—George Melcher, Assistant Superintendent in charge of Research and Efficiency, has favored us with two of his circular letters to principals and teachers and with some data on penmanship work which the Bureau has been doing. Because these communications contain suggestions which may be of service to the other members, they are described in this department.

Members of the Association will probably remember that the Kansas City Bureau some time ago arranged a scale for the measurement of handwriting. The work was done by the same general method that Thorndike used, and the resulting scale is printed in the same form. It includes, however, grade standards and suggestions on the handwriting scale, which make it particularly valuable to the principals and teachers of the local schools. The Bureau now issues a set of fifty specimens of handwriting which have been graded by 89 judges. A median value, as determined by these ratings, has been assigned to each specimen. The specimens are designed to help teachers, principals, and supervisors in standardizing their own scoring of quality of handwriting.

While smaller systems may not wish to undertake as comprehensive a scheme of training for their teachers, yet even the smallest might well consider the advisability of accumulating a set of samples well distributed in quality and scored often enough to determine well established values. These would assist materially in the training of teachers in the use of the scale.

One of the two bulletins mentioned above presents date on overageness. The following points seem worthy of note: In 1896 the percent of overageness in the Kansas City schools was 57, while in 1921 it was 27.5. Since in 1914 the percent of overageness was 45, the rate of reduction in the last seven years has been more than three times as rapid as in the preceding eighteen years. In 1896, 35.6 percent of the children who entered the schools were graduating from the elementary grades; in 1914, 55.2 percent and in 1921, 81.8 percent. The question is raised whether this reduction in overageness will not lower the quality of school work. The answer is made that standardized achievement tests during the past few years have shown steady advancement while the reduction in overageness has been going on.

Many of our members will doubtless be interested in making a similar comparison in their own schools. Your secretary will appreciate it greatly if the findings are forwarded to him for use in this department.

The other bulletin deals with reports on the Courtis Arithmetic Tests. The report shows that the city-wide medians were practically the same in May 1921 as in May 1920. Of all the classes of the city 42.5 percent scored below normal. The following goal is set: "Every class should aim at the standard. Slow classes should reach the lowest normal score in each grade, average classes should reach the standard, and strong classes may score somewhat above standard. However, it is advised that classes should not exceed the upper normal scores in any grade." The upper and lower normal scores referred to are upper and lower quartiles. The author also calls the attention of all teachers and principals to the fact that one school which had always been in the lowest quartile until last year had brought all grades to normal or above. "This splendid piece of work was done in one year under only fair working conditions and with average pupils."

Cannot our readers furnish us with similar records?

University of Michigan.—Last spring, the University of Michigan, Bureau of Tests and Measurements, sent out a bulletin of inquiry to the school people of the state asking for guidance in regard to testing programs. The tabulation of the replies gives an interesting index of the preference which school people in that state are exhibiting in this field. The tabulation follows:

Of 47	replies,	46 wish to include an intelligence test.
" 42	"	25 wish to include at least Grades III to VIII in the program; 32 wish to include six or more of the school grades.
" 31	"	29 wish to include three or more school subjects, and 24 wish to include four or more, though sometimes these four include the intelligence test.
" 31	"	22 wish to use the Courtis Arithmetic Test.
" 33	"	19 wish to use the Monroe Silent Reading Test.
" 47	"	24 wish to use the Ayres Spelling Scale.
" 29	"	19 wish to use the Ayres Handwriting Scale.
" 27	"	19 wish to use the National Intelligence Test.
" 47	"	29 are willing to leave it to the Bureau to decide the order in which the tests shall be given.
" 47	"	44 agree to forward results to the Bureau.

On the basis of this information, the Bureau arranged a program for work in grades three to eight which included courtis Supervisory Arithmetic Tests, Monroe Silent Reading Tests, Buckingham Extension of the Ayres Spelling Scale and the National Intelligence Tests.

We shall look forward with interest to the report on this program and we shall welcome reports of similar programs from other Directors.